NASA/TP–20250003243

Flight Evaluation of In-Flight Strategic Path Planning Automation for Future High-Density Operations

Bryan A. Barrows
NASA Langley Research Center, Hampton, Virginia

Stewart L. Nelson
NASA Langley Research Center, Hampton, Virginia

Tyler D. Fettrow
NASA Langley Research Center, Hampton, Virginia

David A. Karr
Analytical Mechanics Associates, Hampton, Virginia

Mark G. Ballin
NASA Langley Research Center, Hampton, Virginia

David J. Wing
NASA Langley Research Center, Hampton, Virginia

July 2025

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI Program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI Program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI Program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*
- E-mail your question to *help@sti.nasa.gov*
- Fax your question to the NASA STI Information Desk at 443-757-5803
- Phone the NASA STI Information Desk at 443-757-5802
- Write to:
  STI Information Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076–1320

NASA/TP–2025–20250003243

Flight Evaluation of In-Flight Strategic Path Planning Automation for Future High-Density Operations

Bryan A. Barrows
NASA Langley Research Center, Hampton, Virginia

Stewart L. Nelson
NASA Langley Research Center, Hampton, Virginia

Tyler D. Fettrow
NASA Langley Research Center, Hampton, Virginia

David A. Karr
Analytical Mechanics Associates, Hampton, Virginia

Mark G. Ballin
NASA Langley Research Center, Hampton, Virginia

David J. Wing
NASA Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

NASA Langley Research Center
Hampton, Virginia 23666

July 2025

Acknowledgments

The authors would like to thank all collaborators that assisted in this work, including members of the Airspace and Traffic Operations Simulation development team: Michael Havlin, Dr. Elaine Blount, James Sturdy, Doug Mielke, Joshua Kaurich, Patrick Johnson, and others; Matthew Underwood for his system engineering contributions; contributors to the development and testing of the System Integration Lab: Brent Pickering and Nathan Perreau; the Integration of Automated Systems (IAS) flight test team: Dr. Adam Yingling, Nancy Baccheschi, Jan Scofield, Andrew Guion, Ethan Williams, Paul Sampson, Gayle Patterson, Abbigail Waddell, Kevin Monk, Daniel Eng, and Ernest Nwajagu; the NASA flight test pilots: Wayne Ringelberg, Scott Howe, David Zahn, Brent Davis, Bill Fell, John Rucci, and Rachel Simmet; and the flight staff and engineers at Lockheed Martin Sikorsky Aircraft.

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320

National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Available electronically at http://www.sti.nasa.gov

Flight Evaluation of In-Flight Strategic Path Planning Automation for Future High-Density Operations

Summary

Advanced Air Mobility (AAM) is a vision for a future air transportation system that serves new aviation markets, relying on a new generation of technologically advanced aircraft. One potential new market is air transportation within metropolitan areas. Urban Air Mobility (UAM) in the long term envisions thousands of simultaneous air operations over urban areas. It is a key element of the broader AAM vision, delivering economic benefits and alleviating ground congestion. UAM introduces significant challenges in the management of flight operations. The Federal Aviation Administration and NASA propose concepts for evolutionary development toward high-density urban operations. Initial low-tempo UAM operations will take place in the current regulatory and operational environment, followed by evolutionary changes to increase operational tempo. As a complement to these near-term activities, this study takes a transformational approach, addressing a key technical challenge in managing future high-density operations.

Human operators may not reside on UAM aircraft in the future. Even if they do, high traffic density and operational complexity may make traditional human operator tasks challenging. AAM operations are therefore anticipated to rely on collaborative and responsible automation. Flight Path Management (FPM) is an automation concept that may be a critical functional enabler in future operations. FPM provides the function of strategic and dynamic flight path planning in the presence of other users sharing the airspace, and in the presence of restrictions and planning constraints necessary in the management of traffic flow.

A flight test of FPM was conducted using a reference prototype FPM automation system developed by NASA. Test goals included a verification of the automation's core functions, exploration of function behaviors in a flight environment, discovery of unknown unknowns, and validation of the operating environment simulation used to develop the concept and technology. The test used two aircraft and a simulated high-traffic urban environment. Testing was performed using helicopters as surrogates for future electric vertical take-off and landing (eVTOL) aircraft. One helicopter was equipped with the reference prototype FPM automation system. Another helicopter was used as a cooperating intruder, sharing its flight intent with the automation-equipped aircraft. The two aircraft engaged in conflict encounters in a live-virtual-constructive operating environment. A model of future high-density urban airspace was developed and populated with up to 330 virtual traffic aircraft that also cooperated by sharing their flight intent.

Flight test results provide positive indications that a vehicle-centric implementation of FPM can be made operational in the future. FPM has the potential to address critical challenges in managing the high-density traffic operations envisioned for long-term UAM operations. All functional performance success criteria for maneuvers in an airspace environment containing some area restrictions were met or exceeded. The research prototype automation technology reliably supported the core functions of intent-based conflict detection, strategic conflict resolution, conflict prevention, and arrival time compliance. Pilot feedback suggests human operators can be in the FPM decision-making loop at high traffic density. Test maneuvers involving operations within flow corridors were less successful, primarily due to having tighter volume constraints within the airspace and fewer degrees of freedom available for conflict resolution maneuvering. A specific trajectory management approach and its enabling automation technology need to be developed to support high-density operations within corridors. An initial simulation validation was performed using flight test data. Results suggest that with appropriate control of host platform capabilities, such as computational power, the existing simulation can be used to identify behavioral trends at both a vehicle level and an airspace system level.

Contents

List of Figures

List of Tables

Nomenclature

4D	four-dimensional (three spatial dimensions and the time dimension)
AAM	Advanced Air Mobility
ADS-B	Automatic Dependent Surveillance–Broadcast
AMM	Autonomy Mission Manager
AOP	Autonomous Operations Planner
BBTG	Behavior-Based Trajectory Generator
CAS	calibrated airspeed
CD	conflict detection
CP	conflict prevention
CPA	closest point of approach
CPU	central processing unit
CR	conflict resolution
DFW	Dallas Fort Worth
ETA	estimated time of arrival
EUTL	Efficient Universal Trajectory Language
eVTOL	electric vertical take-off and landing
FATCAT	FPM/AOP TPUBs Calibration Analysis Tool
FMS	flight management system
FPM	Flight Path Management
GCS	Ground Control Station
HITL	human-in-the-loop
IAS	Integration of Automated Systems Subproject
IAS-1	IAS phase 1
IFTS	Integrated Flight Test System
LOS	loss of separation
LVC	live-virtual-constructive
MOE	measure of effectiveness
MOP	measure of performance
OPV	Optionally Piloted Vehicle
PBGA	Pattern-Based Genetic Algorithm
PSU	Provider of Services to UAM
RTA	required time of arrival
SARA	Sikorsky Autonomy Research Aircraft
SICR	strategic intent-based conflict resolution
SUA	special use airspace
TPUBs	Trajectory Prediction Uncertainty Bounds
TRL	Technology Readiness Level
TTFLOS	time to first LOS
UAM	Urban Air Mobility
UI	user interface
UML-4	UAM Maturity Level 4

1 Introduction

Advanced Air Mobility (AAM) is a vision for the future of air transportation that aims to serve emerging aviation markets through a new generation of technologically advanced aircraft (Ref. 1). One such market is metropolitan air transport, referred to as Urban Air Mobility (UAM). As a central component of the broader AAM vision, UAM anticipates thousands of simultaneous air operations over urban areas in the long term. It offers economic benefits through an efficient means of transportation by alleviating ground congestion.

UAM introduces significant challenges in the management of flight operations, such as high-density airspace management, integration, and regulatory compliance, in unison with existing airspace operations. The Federal Aviation Administration and NASA have proposed concepts for evolutionary development toward high-density urban operations. Initial low-tempo UAM operations will take place in the current regulatory and operational environment, followed by evolutionary changes to increase operational tempo (Ref. 2). As a complement to these near-term activities, this study takes a transformational approach, investigating the technical challenge of in-flight path planning that achieves traffic separation while complying with airspace restrictions and operational constraints in tandem with existing traffic patterns. Safety and efficiency of increasingly complex AAM operations are expected to require extensive use of automation, ranging from controlling revolutionary new aircraft types to managing flights dynamically in high-tempo airspace and aerodrome operations. Airborne and ground automation will play central roles in assisting AAM operators with managing the flight paths of their fleets. The NASA AAM Project was established to accelerate development and operational adoption of the vision. Project goals include the exploration and development of key automation capabilities. Flight Path Management (FPM) is one such capability potentially needed for the high-density operations envisioned for AAM at maturity. FPM provides strategic and dynamic flight path planning in the presence of other users sharing the airspace, and in the presence of restrictions and planning constraints necessary in the management of traffic flow. Arguably, the most important functions of FPM are detecting and resolving conflicts. A conflict is a predicted loss of separation (LOS) with another aircraft. As a planning function, FPM performs strategic deconfliction, maintaining mission objectives as it maintains separation.

UAM Maturity Level 4 (UML-4) is roughly equivalent to the FAA's mature operational stage (Ref. 2). It is defined as having air transportation "accessible and attractive to a significant percentage of the public for travel between high-density, origin-destination pairs (e.g., commercial airport to business district)" (Ref. 3). UML-4 is envisioned to have hundreds of simultaneous aircraft aloft over the served urban area in all weather conditions, resulting in an extremely high level of traffic density compared to current operations. Air traffic management functions that ensure safe aircraft separation, maintain high traffic flow, and accommodate operator preferences cannot support UML-4 in their current state. These functions rely heavily on traditional air traffic management methods, such as human situational awareness and decision-making, which may lack efficiency and scalability given the increased operational tempo and traffic density. FPM automation may be a critical enabler of UML-4 operations. A detailed description of the FPM intended function is provided in Reference 4.

Decision makers need technology advancements and results from foundational and operations research to inform the ambitious AAM vision. Concepts similar to FPM have been investigated for airline operations for more than two decades, and their enabling technologies have been prototyped for research. Most of the research has been performed using simulated air traffic environments. The next logical phase of research and development is to perform evaluations in flight to verify expected performance and to discover unknown challenges. In this study, a reference prototype FPM automation system was developed and evaluated in flight for anticipated future operations at UML-4.

A flight test was conducted through a joint undertaking of the NASA AAM Project's Automated Flight and Contingency Management and National Campaign Integration of Automated Systems (IAS) Subprojects. Referred to as IAS phase 1 (IAS-1), the flight test was conducted with Lockheed Martin Sikorsky aircraft. The test was made up of a series of campaigns and was completed in October 2023.

Testing was performed using two helicopters as surrogates for future electric vertical take-off and landing (eVTOL) aircraft. One helicopter was equipped with the reference prototype FPM automation system. Another helicopter was used as a cooperating intruder, sharing its flight intent with the FPM automation-equipped aircraft. The two aircraft engaged in conflict encounters in a live-virtual-constructive (LVC) UML-4 operating environment. A model of future UML-4 airspace was developed and populated with up to 330 virtual traffic aircraft that also cooperated by sharing their flight intent.

This paper describes the goals, approach, and conduct of the flight test, and presents results and insights gained. An overview of FPM is provided, followed by an overview of the developed reference FPM automation. The research objectives, scope, and approach are then described. The flight test system and the an LVC test environment are presented with a level of detail appropriate to the research focus. Test design and execution information is provided. The paper then presents results and analysis, and it concludes with a summary of insights gained and recommendations for future research. Appendices provide additional detail.

2 Motivation

This research addresses foundational concept feasibility and system architecture design issues that must be resolved to realize the mature UAM vision. For system implementation to be affordable and timely, these issues should be addressed before large-scale investments in system infrastructure are made. All functions and architectural alternatives should be explored. FPM is one function of many that will be needed to realize the vision, and the scope of the current research is focused on one of several potential architectures. The specific solution addressed is onboard self-management and a vehicle-centric distributed architecture. While carrying significant challenges associated with transformational change, the addressed solution also offers transformational benefits. Results should therefore provide important contributions to the body of knowledge needed for long-term implementation decisions.

UAM operating at UML-4 is an extremely challenging use case that demands focus on technical challenges associated with operations at very high traffic density levels. An increased confidence in feasibility and viability of UML-4 operations can provide a mature-state anchor for National Airspace System modernization. Such an anchor may also provide guidance for near-term modernization efforts. Therefore, steps taken toward enabling UAM will pay dividends in advancing operations in other operating domains, and by extension, advancing benefits for all operators and airspace systems.

In December 2023, RTCA, Inc. announced the Member Report on Digital Flight (Ref. 5). The new report seeks to:

> . . . develop a framework of aircraft and system requirements that would be needed to support future aviation digital flight operations. UAM and AAM envision a safe and efficient aviation transportation system using highly automated aircraft to operate and transport passengers or cargo at lower altitudes within urban, suburban, and non-urban environments. The purpose of this forum is to develop industry consensus around the required capabilities needed to ensure the integration of new entrants and legacy operations in a safe and efficient manner.

The establishment of this forum as well as other committees and working groups illustrate the need for foundational operations research and technology development to support these ambitious visions for future flight operations.

Community needs include the establishment of design alternatives for airspace system architectures, system infrastructure requirements, performance standards for all entities operating within the system and the system itself, and requirements and procedures related to flight operations. The current research activity provides data and information to support these needs. Intended audiences are industry, which will have the task of developing FPM automation systems; the FAA, to be tasked with regulating FPM automation systems and operations; standards bodies, tasked with advising the FAA on means of compliance; future operators considering use of FPM automation systems; program managers funding UAM/AAM research and development; and the research community, tasked with envisioning and developing concepts and technology related to future aviation operations.

The activity is an initiating element of research focused on advanced at-scale operations envisioned by the UAM/AAM community. As a first activity, its scope is limited. Within its scope, the activity seeks to explore the performance and operational behaviors of the FPM function, making use of a representative automation technology. Performance of various critical technology functions are examined to determine operational behavior in a relevant environment. The research seeks to establish performance benchmarks that will be useful in establishing future functional requirements. It will also inform needs for further technology development or refinement.

One of the most important needs in airspace management research is validated air traffic operations simulations. Due to the expense of multi-aircraft flight testing, simulations are critical to the development, refinement, and validation of the envisioned future operations at the UML-4 level and above. While a balanced study of FPM in a UML-4 operational environment requires both simulation and flight testing, air traffic simulation is an irreplaceable tool.

- Simulation is the only cost-effective way to perform the bulk of needed FPM research. The UML-4 environment is expected to have hundreds of participating aircraft in proximity to each other, and the number of design and operating parameters associated with system concepts such as FPM are extensive.

- Simulation enables experimental control for internal validity. All variables are controlled to the extent possible except for the manipulated variable, thereby providing high confidence that measured results are caused by the manipulation.

- If internal validity is high, studies can be designed to minimize nuisance variation, which makes it easier to distinguish desired signal from undesired noise. For instance, data links can be modeled as perfect, thereby making it easier to study other input/output relationships.

- System functional requirements and their rationale can be determined, assessed, and validated by varying assumptions in simulation.

Flight testing is also critical because it enables air traffic simulation results to be trusted.

- Testing in conditions as close as possible to the actual environment increases confidence that study results are likely to replicate in the field (external validity).

- Simulation may rely on models that are simplified in a way that produces misleading results, especially if models are used outside the bounds of known validity. Flight testing can build confidence that simulation results that rely on necessary extrapolations are valid.

- Flight tests can provide data supporting the improvement of simulation models.
- Flight tests may discover unknown unknowns, representative of undiscoverable factors in controlled simulations.

As the initial flight test of FPM capabilities, the current activity is a critical element of UML-4 research.

3 The Flight Path Management Automation Concept

UAM is among the most complex operating environments envisioned for commercial aviation. UAM viability will depend on achieving many technical breakthroughs, including possibly introducing a new operating mode that automates traffic separation under exceedingly challenging and diverse constraints. Examples of UAM constraints that must be accounted for include:

- Limited endurance of eVTOL aircraft, requiring ultra-efficient flight paths.
- Reduced separation standards to enable sufficient traffic density and volume.
- Dynamic scheduling of high-demand vertiports, requiring reliable on-time arrival.
- Area constraints defined by municipalities, air traffic control, and airport traffic flows.
- Altitude and wind constraints defined by buildings, terrain, air traffic control, and airport traffic flows.
- Coordination procedures that consume decision-making time.
- Wide-ranging aircraft performance of many UAM aircraft configurations.
- Interactions with traffic operating under different flight rules.

An approach is needed that can address all these constraints simultaneously. Many of these constraints are time-dependent and may change during the course of a flight. A sequential approach, such as use of a strategically deconflicted pre-departure plan followed by an in-flight tactical "detect and avoid" capability, is anticipated to be insufficient to provide the required reliability and robustness for sustainable commercial operations. FPM is a proposed function that considers all constraints simultaneously by strategically replanning the remaining portion of a flight each time a constraint is changed or a new constraint is encountered. FPM was first defined in Reference 6 in an effort to identify the various functions that are performed in managing the airborne phase of flight, whether performed by human operators, automation, or a combination of humans and automation.

FPM provides the function of strategic and dynamic flight path planning in the presence of other users sharing the airspace, and in the presence of restrictions and other planning constraints. The term 'strategic' in this context is defined as planning that achieves mission objectives and retains them when replanning is necessary. Strategic planning can be contrasted with tactical planning, which suspends mission objectives to achieve a more important objective, such as maintaining safety. FPM is envisioned to work in concert with a tactical planning system, which provides an additional layer of safety. A detailed description of the FPM intended function is provided in Reference 4.

Mature state AAM concepts of operation will be strongly impacted by the distribution of system functions and the allocation of responsibility and authority to agents within the system. Design alternatives for the concepts and their supporting system infrastructures are the subjects of ongoing research. Evaluation of these alternatives will depend on the study of system component functions and architectural alternatives involving those functions. Locating the FPM functional capability and its enabling technology with each participating aircraft is one such architectural alternative and is the architecture assumed for the flight test. Each aircraft performs the FPM function while sharing information relevant to other system participants and receiving information relevant to itself from other participants.

For such a distributed-agent architecture applied to future UML-4 operations:

- The vehicle operator is responsible for separation from traffic while conforming to area constraints such as corridors and restricted airspace regions.

- Schedule and airspace constraints are issued from external services. The vehicle operator is responsible for complying with them, even when they change in flight.

- The operator employs collaborative and responsible automation to perform the separation function. Collaborative and responsible automation is defined as automation "assured to perform specified functions such that human monitoring and mitigation of potential failures of those functions is no longer necessary" (Ref. 3).

- Data exchange and any necessary coordination is accomplished through airborne and ground-based data links.

4 FPM Automation Technology

Because of high traffic density in future UML-4 operations, traffic management issues may arise rapidly. Strategic path planning solutions may be needed to maintain system predictability. Fast and precise responses may be required to maintain system stability, robustness, and possibly resilience. For these reasons, the study assumed automation will be an enabling system component.

Human operators will likely have a role in the overall FPM function. Operators may be system users or service providers, and they may be located onboard or offboard the aircraft. The role of the human operator is beyond current scope and will require extensive research. For the current activity, automation is assumed to provide alerting and trajectory change advisories to a human operator residing on each aircraft.

4.1 Intended Function

An automation system performing dynamic path planning generates operational intent to determine a four-dimensional (4D) flight path, also known as a 4D trajectory, which specifies a predicted latitude, longitude, and altitude of an aircraft at each future time point. The automation system performs the following tasks.

- Create the flight path.

- Monitor the flight path and the factors that may impact it.

- Evaluate ongoing acceptability of the flight path and proposed changes.

- Revise the flight path, as needed, to sustain desired qualities.

• Coordinate the flight path with other airspace users and service providers.

Five principal qualities are desired for all trajectories created by the automation system: feasibility, deconfliction, harmonization, flexibility, and optimality (Refs. 4, 7).

1. A feasible path is one that conforms to the aircraft performance and range capabilities; complies with the airspace structure, rules, and constraints; avoids the terrain and charted obstacles; and meets the arrival constraints.
2. A deconflicted path is one that avoids unsafe proximity to known aircraft, dynamic obstacles, inclement weather, and other emergent airspace hazards.
3. A harmonized path is one that follows cooperative rules and procedures to ensure that the use of the airspace is coordinated with other airspace users.
4. A flexible path is one that provides adequate maneuvering room to ensure future flight path changes, if needed, are available and feasible.
5. An optimal path is one that best achieves the operator's business objectives for the specific flight.

The computed flight path must account for mission objectives, aircraft performance, traffic flight information, current and predicted atmospheric conditions, obstacles and terrain constraints, community rules, and variations in airspace and aerodrome configurations. It must be interoperable with other systems, such as those that support the tactical hazard avoidance function. If the FPM automation system is providing decision support to a human operator, it must also account for pilot inputs and provide contextual display information, flight conformance status, system health and alerts, and contingency response options.

4.2 Reference Implementation

A prototype FPM automation system developed by NASA was used for the flight test. This prototype, a modified version of the NASA Autonomous Operations Planner (AOP), has been developed and tested using batch and human-in-the-loop (HITL) simulations. AOP is designed so that each running instance of the software performs dynamic path planning for one aircraft known as the ownship. In this concept of FPM automation, many aircraft may simultaneously be equipped in this way, each one serving as the ownship to its own instance of AOP.

AOP is a research prototype automation system and is therefore only intended as a reference capability. As noted, the UML-4 maturity level describes automation as "collaborative and responsible." FPM automation may be required to be responsible for performing functions without human backup in a future system that is certified for operation. As a research prototype, AOP was not designed to meet the level of assurance required to be considered collaborative and responsible, but it provided the sophisticated FPM functionality required by this research, while flight operations were protected by the fault tolerant architecture of the test aircraft and by safety mitigation procedures.

Development of AOP began in the late 1990s to support the exploration of distributed air/ground traffic management concepts (Ref. 8). These concepts had the goal of increasing capacity of the National Airspace System, and were based on the following propositions:

• An inability to predict future airspace system states with high accuracy requires a management approach capable of timely response to unanticipated changes in the environment or situation. Responses must not cause a degradation in stable system operation.

- A distributed system architecture relieves workload bottlenecks associated with centralized architectures that rely on human operators.
- Placing a system management function in the location that contains the most timely and highest-quality information produces a highly responsive system and a reduced need to exchange information between management functions and execution functions. Prompt responses enable operators to implement solutions when they are still applicable and before problems in the airspace become worse.

Reference 9 provides further detail of the distributed air/ground concept of operations for which AOP was designed.

AOP was originally designed for research into the the application of aircraft autonomy to commercial transport operations, primarily for airlines. It was designed to be an integrated suite of strategic and tactical airborne trajectory management functions compatible with commercial transport avionics architectures and industry standards. AOP was modified for UAM operations that employ eVTOL aircraft, as detailed in Section 4.3.

AOP capabilities deemed ready for flight testing included intent-based conflict detection (CD), which predicts future LOS, strategic intent-based conflict resolution (SICR), which produces conflict-free trajectories as advisories, and the Behavior-Based Trajectory Generator (BBTG), which computes detailed 4D trajectories to support the other two capabilities. In this report, conflict resolution (CR) refers to SICR, since other CR capabilities of AOP were not considered.

These capabilities align with the FPM automation system's objective to maintain five flight path qualities as follows:

1. Feasibility
    - Trajectories generated by the BBTG use a model of the ownship's performance capabilities and operating limits.
    - The BBTG is aware of the hosting aircraft's state, including energy on board.
    - Generated trajectories comply with crossing restrictions, special use airspace (SUA) restrictions, and instructions to arrive at a required time.
    - Some constraints can be relaxed to satisfy higher-priority constraints.
2. Deconfliction
    - Deconflicted trajectories produced by AOP's CR avoid traffic, current and forecast 4D convective weather, and 4D restricted airspace.
    - Multiple CR alternatives are identified to meet replanning objectives, as available. Alternatives are offered for lateral path changes, vertical path changes, a combination of lateral and vertical changes, and speed changes.
3. Harmonization
    - AOP's CR maneuvers are coordinated with traffic using priority rules (right-of-way rules).
    - Resolution maneuvers employ conflict prevention (CP): computed resolutions will not create new conflicts with other traffic.
    - Resolution computations respect instructions to arrive at a required time and respond to changes in required arrival time instructions.

4. Flexibility

    • Flexibility was not in scope for this effort. In previous research, AOP was used to demonstrate the concept of flexibility preservation, but the capability was not retained in the AOP baseline.

5. Optimality

    • AOP's CR advisories provide paths of lowest energy, shortest distance, or shortest time (depending on operator preferences).

These capabilities are necessary, but probably not sufficient, for full UML-4 operations. Two goals of the flight test are to validate the need for the existing capabilities and to identify additional needed functional capabilities. Additional capabilities have been envisioned for each of the five automation qualities and will be the subjects of future research and development. References 7 and 10 describe the flight-tested capabilities of AOP in detail and provide information about additional capabilities currently under development.

Primary components of one instance of AOP are shown in Figure 1. The AOP Conflict Detector, which

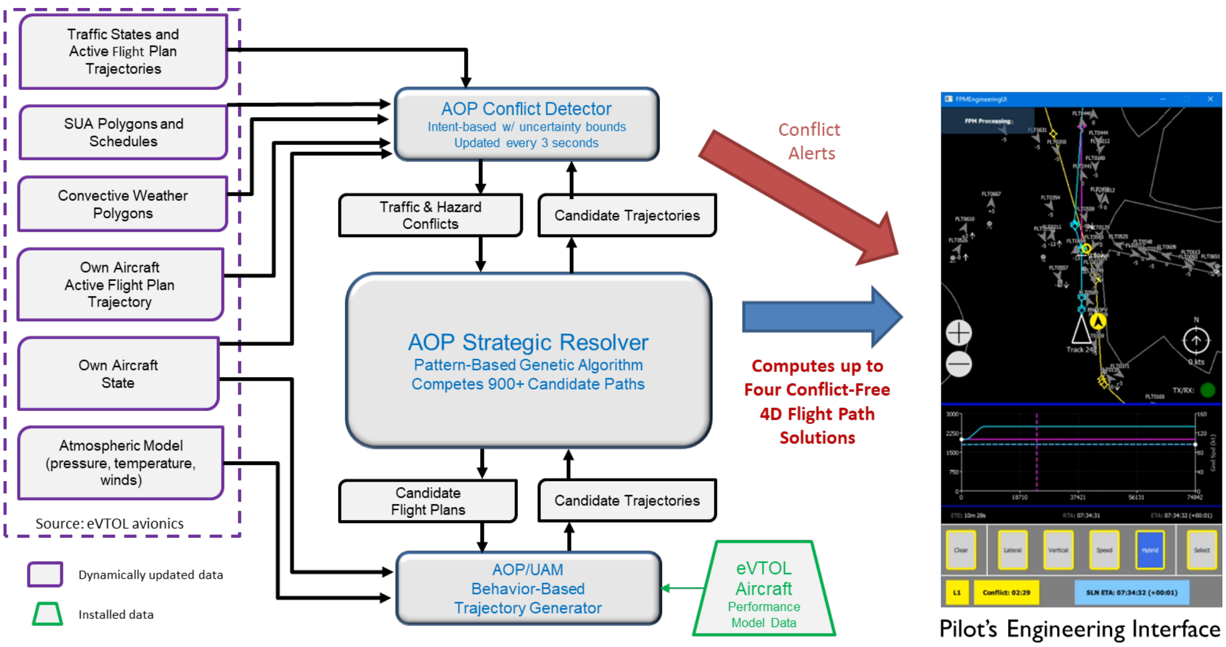


Figure 1. Functional architecture of AOP for the flight test.

implements the CD function, identifies traffic separation conflicts, boundary crossings into restricted airspace or hazardous weather regions, and noncompliance with trajectory constraints. The CD function propagates active trajectories of all aircraft forward in time, using intended trajectories when they are known. The predicted 4D trajectory of the ownship is the *ownship trajectory*. The predicted 4D trajectory of each other aircraft is a *traffic trajectory*.

Trajectory Prediction Uncertainty Bounds (TPUBs) applied to each segment of the 4D trajectory of the ownship and of each traffic aircraft are an innovative technique in AOP that accounts for flight guidance errors or navigational errors (Ref. 11). TPUBs specify maximum anticipated deviations from the predicted 4D trajectory as the aircraft follows its flight path. These deviations can be specified independently in various directions: above the path, below the path, laterally to the right or left of the path, and along the path. The CD function in AOP performs a worst-case analysis on these deviations to determine whether aircraft separation requirements may be violated or unwanted airspace penetrations may occur during the actual flight. The effect is to add margins of safety to the zones of separation protection around the aircraft, tailored to the kinds of prediction errors that are likely to occur.

Priority (right-of-way) rules are incorporated into logic that alerts the pilot of a conflict to be resolved. The aircraft that is burdened under these rules is the first aircraft to become responsible to resolve the conflict. When AOP's priority rules determine that the ownship is responsible for resolving the conflict, the process of SICR is triggered in the AOP Strategic Resolver, shown in the center of Figure 1. The instance of AOP for the non-burdened aircraft executes SICR only if the burdened aircraft does not compute and propagate a conflict-free trajectory within a certain time interval.

The SICR function generates a set of candidate modifications to the active flight path to resolve conflicts and other issues identified by the CD function. The function uses a genetic algorithm called the Pattern-Based Genetic Algorithm (PBGA) to perform an iterative search for solutions using a defined set of maneuver patterns (parameterized trajectory modifications such as a path stretch or offset maneuver) for the resolution candidates. Several patterns are available, classified into four groups according to the degrees of freedom controlled by each pattern: lateral resolutions, vertical resolutions, a hybrid combination of lateral and vertical resolutions, and speed-change-only resolutions. As configured for the flight test, PBGA iterates over 20 generations, each of which has a population of 20 trajectory candidates, for each degree-of-freedom class. A fitness function ranks candidates based on specified optimization parameters. The SICR function uses the CD algorithm to prevent acceptance of any trajectory candidates that either do not resolve the identified issue or create new issues.

The specific 4D path of each trajectory candidate is created by the component labeled AOP/UAM Behavior-Based Trajectory Generator in Figure 1. The BBTG uses a model of the hosting aircraft (eVTOL Aircraft Performance Model Data in the figure) that incorporates performance and procedural constraints to ensure each trajectory created can be flown by the hosting aircraft. Data are provided to these functions in near-real-time, as shown by the blocks on the left side of Figure 1.

Several other CD and CR functions are available in AOP but were not used in the flight test and are not shown in Figure 1.

4.3 Modifications for UAM UML-4 Research

Modifications and enhancements were made to AOP software to support the flight test (Ref. 10). These modifications and enhancements were incorporated into the baseline software so that they may be used in future work when enabled by the AOP configuration files.

Additional configuration settings based on concept assumptions for the IAS-1 flight test, such as UML-4 separation standards, were applied to further modify the behavior of AOP. Multiple configurations of AOP were created in order to vary the time parameters for CD, conflict alerting, and CR as described in Section 7.

- Many algorithmic efficiency improvements were made to accommodate UML-4 traffic density levels, which far exceeded traffic densities studied in previous airline transport research.

- Performance models of two reference eVTOL vehicles were incorporated. Performance models were based on the all-electric variants of the "Quadrotor" and "Lift+Cruise VTOL Aircraft" vehicles of Reference 12. The performance models were used by AOP to generate trajectories representative of future UAM aircraft within the hosting aircraft's flight envelope.

- Models of battery energy storage and use for the eVTOL vehicles were incorporated in the BBTG. Minimization of battery energy usage was added to the fitness function in PBGA as a trajectory-optimizing measure.

- The speed-change-only degree of freedom for strategic replanning was implemented in PBGA.

- Separation standards assumed (but not validated) to be appropriate for UML-4 were incorporated in the configuration of AOP for UAM. The standards used were 1500 feet lateral separation and 450 ft vertical separation, roughly equivalent to RTCA detect-and-avoid well clear volumes (Ref. 13).

- Lookahead and alerting time horizons for UML-4 were modified based on engineering judgment and preliminary batch testing. Batch testing is described in Section 5.2.

- Trajectory prediction functions were modified to accept ground-referenced 4D trajectory intent shared between aircraft. The data exchange format used was the Efficient Universal Trajectory Language (EUTL), described in Reference 14.

- AOP's tactical conflict management functions were disabled. The tactical hazard avoidance function was not a subject of study in the current activity.

- AOP functions for all trajectories not associated with an aircraft's auto-flight system coupled lateral and vertical navigation (LNAV/VNAV) modes were disabled.

- AOP's TPUBs values were set based on discovery of the 4D trajectory conformance capabilities of the two test aircraft as described below.

Due to the assumptions of 4D trajectory conformance in the UAM domain and the measured trajectory-conformance capabilities of the test aircraft, the TPUBs values in each dimension were held constant throughout the flight path. Sophisticated features of TPUBs such as applying a continuously increasing along-path uncertainty on each trajectory segment to account for uncertainty in the mean groundspeed of the aircraft were deemed inapplicable. Other ways to "fine-tune" TPUBs were considered out of scope for this effort due to limited opportunities to test actual aircraft performance.

The 4D trajectory conformance of Sikorsky Autonomy Research Aircraft (SARA) were assessed in preliminary test flights. Flight guidance was tasked with following EUTL flight plans constructed for this purpose (not created by AOP). After an examination of the conformance of the guidance to the flight plans using techniques described in Appendix E, the TPUBs for the flight test were set to ± 150 feet lateral, ± 25 feet vertical, and ± 3 seconds along path at all points along the 4D flight plan (Ref. 10), meaning that the aircraft was predicted to pass each point along the path of the 4D flight plan within 150 feet horizontal distance left or right of the predicted latitude and longitude, 25 feet above or below the predicted altitude, and 3 seconds after or before the estimated time of arrival (ETA) at that point. AOP therefore assumed during the flight test that each aircraft would stay within these bounds continuously.

The selection of these particular values of TPUBs was based on engineering judgments regarding the limited opportunities for preliminary assessment of trajectory conformance and the acceptable levels of prediction uncertainty for the flight test, and must not be construed to represent either capabilities or limitations of the test aircraft guidance systems.

Modifications of AOP are described further in Reference 10.

5 Research Approach

An overarching goal of the sponsoring project was to determine the level of maturity of the concept and its enabling technology. To assess maturity level, performance of AOP functions relevant to UAM were examined in a ”relevant environment”, as specified for NASA Technology Readiness Level (TRL) 5 (Ref. 15). The flight test activity also sought to gain insights regarding behavior of future UML-4 operations at the airspace system level, representing integration with wide-range airspace operations, in addition to the focus of vehicle-function behaviors.

5.1 Scope

The reference FPM automation system was evaluated in flight in representative UAM missions under simulated UML-4 conditions. A live-virtual-constructive (LVC) approach was used. Helicopters were used as surrogates for future eVTOL aircraft.

One helicopter, referred to as the ownship, was equipped with the reference FPM automation system. This helicopter was the only aircraft in this research activity capable of receiving plans from FPM automation. The other helicopter, designated as the intruder, was used to create encounters with the ownship. The intruder shared its current state and trajectory intent with the ownship. To simulate a high-density airspace environment representative of UML-4, hundreds of virtual background traffic aircraft were introduced. These aircraft, whose movements were prerecorded from traffic simulations using the same UML-4 airspace conditions, also shared their state and intent with the ownship. Intent sharing is a fundamental component of a cooperating operating environment, as envisioned by the UAM concept. The ownship performed FPM functions in order to remain safely separated from all aircraft, whether virtual or real. In order to isolate ownship functional beahvior in the flight test data, the intruder purposefully did not react to the ownship’s maneuvers. All traffic were in their climb, cruise, or descent phases of flight.

Several important topics of research are beyond the scope of the current activity. Beyond-scope topics include:

- Investigations of the ownship and traffic in takeoff, in transition from vertical to translational flight after takeoff, in transition from translational to vertical flight prior to landing, and in landing phases of flight
- Collaborative and responsible automation expectations or requirements development for FPM enabling technology
- Flow management technologies supporting capacity-constrained operations within corridors and airspace in immediate proximity to vertiports
- Formal investigations of pilot decision aiding by the reference automation’s advisories
- Development of human/automation teaming practices and procedures
- Formal human factors assessments of technology, procedures, or concept feasibility
- Impacts of coordination delays involved in flight path changes, such as potentially mandatory coordination with a ground-based Provider of Services to UAM (PSU) prior to a maneuver
- Interoperability of FPM with tactical hazard avoidance functions

- Sharing airspace with nonparticipating vehicles and coordinating with their traffic management service providers

- Closed-loop dynamic interactions between FPM functions and constraint-setting flow management functions (e.g., metering schedulers).

5.2 Previous Research Activities

Two simulation studies were conducted in preparation for the flight test activity. Both studies used batch traffic simulations for baselining and improving FPM performance in a UML-4 operational environment. The second study also conducted near-real-time HITL simulations for one piloted aircraft in a simulated UML-4 traffic environment. All simulated aircraft performed FPM functions, with a full instantiation of AOP operating on each simulated aircraft. A representative UML-4 operational environment was developed for the studies, and is described in Section 6.2.2. More than 150 traffic aircraft operated simultaneously within the defined airspace region, with vehicle routing distributed across 50 vertiport locations. System-level performance was defined to include measures of the FPM automation's ability to meet mission objectives while maintaining operational safety.

The first study was referred to as FPM-1, and is detailed in Reference 16. A primary challenge in FPM-1 was the lack of a reference case for automated operations in a UML-4 operational environment. The study's primary goal was to establish a reference baseline for nominal en route flight operations, with an emphasis on the feasibility, deconfliction, and harmonization qualities defined for FPM. Objectives included understanding and baselining the FPM automation's performance for use in future research and development activities, and discovering unknown behaviors. A total of 72 scenarios, encompassing over 50,000 flights, were executed to understand the impacts of various path deconfliction approaches, vehicle maneuvers, and schedule contingencies. Dozens of scenario parameters variations were tested. TPUBs availability, traffic intent availability, required time of arrival (RTA) delay, wind error, maneuverability, and traffic demand were among a set deemed of high importance throughout the study. At a system level, LOS and RTA conformance were two metrics in the study that were shown to have the largest impact. The findings from FPM-1 provided a solid foundation for the second study, FPM-2. The FPM-1 performance evaluation also provided confidence in system readiness for conducting the FPM-2 HITL activity.

The goals of the FPM-2 study were to conduct further testing and verification of the reference automation functions, inform the design and selection of flight test scenarios, and evaluate an engineering user interface (also known as the engineering UI) developed for the flight test. Results of FPM-2 were used extensively to refine the reference automation and the engineering UI. Results are documented in an internal FPM-2 validation report. Reference 17 provides additional FPM-2 information.

5.3 Assumptions

A detailed concept of operations for high traffic density UAM operations and services was not available at the time of this flight test. The following assumptions are made for the current research.

5.3.1 FPM Assumptions

- FPM automation detects conflicts, future incursions of restricted airspace, or noncompliance with time constraints associated with the hosting aircraft's active trajectory and provides trajectory change resolution advisories to a human operator onboard the aircraft. The operator evaluates the advisory options, selects one of them, and assigns the selection as the new active trajectory for the hosting aircraft. Creating a new active trajectory in this way is referred to as executing the resolution.

- All aircraft conform to ground-referenced 4D trajectories with reliable precision.
- A representative future PSU provides schedule constraints to the FPM subject aircraft. The PSU provides information services but not deconfliction.
- Aircraft may change their 4D trajectories at any time and for any reason, but are required to share their new operational intent in a manner enabling all other system participants to have full awareness of the changed flight plan. Changed flight plans are shared at the time of the change.
- All aircraft are FPM-capable at a UML-4 level and are participating in FPM operations. Mixed equipage operations are not currently considered.
- Priority rules (also known as right-of-way rules) are nominally in effect, but because the ownship is the only aircraft in the test with FPM capability, the ownship yields right of way to all other vehicles. For example, during encounters with other aircraft, the ownship maneuvers to resolve conflicts rather than waiting for the other aircraft to maneuver. Appendix G gives the full details of the assumed priority rules, which were also used in the FPM-1 and FPM-2 studies.
- Traffic departing from vertiports would normally be required to perform strategic CD and CR before departing. Recorded background traffic cannot do this. Recorded traffic shares intent prior to departure in sufficient time for the ownship FPM functions to detect and resolve any conflict.
- Assumed UML-4 separation standards are 1500 ft lateral separation and 450 ft vertical separation. These are roughly equivalent to RTCA DO365 DAA "well clear" standards (Ref. 13), and serve as a representative placeholder.
- Use of the "well clear" standards for the separation function assume that a determination of priority between strategic separation and hazard avoidance is based solely on time before LOS. At 30 seconds prior to predicted LOS, hazard avoidance functions are assumed to override strategic separation functions.
- Time is required to allow for system delays or decision-making before executing flight path change advisories provided by the FPM function. Time durations are assumed to be on the order of a few seconds.

5.3.2 Airspace Assumptions and Simplifications

- No hemispherical rules or similar rules are used to constrain altitude or direction of flight.
- There are no equivalents to standard terminal arrival routes or standard instrument departures in the represented UML-4 airspace.
- No routes have assigned speeds or directions.
- Aircraft scenarios begin with flight plans that have been developed in the represented airspace to a) provide a direct route between departure and destination vertiports, or b) if a direct route is not available due to obstructions such as SUA, routes that include defined waypoints. Aircraft that have

not deviated from their initial flight plan are on these routes, and as a result, scenarios may begin with multiple aircraft in trail on these routes.

- Aircraft are free to deviate laterally and vertically, using any part of the airspace not within an SUA region. When changing their flight plans, there is no requirement for aircraft to conform to a specific altitude or navigation fix. Aircraft may also change speed without restriction to any specific speed or speed increment.

5.3.3 Test Assumptions and Simplifications

- The PSU function passes state and intent information from the intruder to the ownship with no modification and no time delay other than data transmission delay.

- Data links used in the flight test have adequate performance for the purpose of the test and are therefore assumed not to interfere with FPM in-flight evaluations.

- The onboard platforms hosting the FPM automation have computational performance adequate to support FPM functions in a future UML-4 environment.

5.4 Research Elements

The flight test activity sought to assess the current state of technology and gain insights leading to further development of the concept, the technology, and the supporting research tools. As complex engineered systems, air traffic management systems exhibit emergent behaviors that defy model-based prediction. The research therefore sought to discover unexpected characteristics and system behaviors. The research approach was made up of four research elements that map to these objectives. The Function Verification element assesses the automation, the Concept and Technology Advancement element provides data to support future development, the Discovery element supports discovery of unknown unknowns, and the Simulation Validation element affirms system behavior across tested scenarios.

5.4.1 Element 1: Function Verification

The function verification research element seeks to determine whether expected performance occurs in a relevant environment. The FPM automation reference implementation was assessed to determine if selected capabilities perform in flight as intended and expected. Assessment was performed using functional performance benchmarks that are expected to be useful in developing future requirements. Automation performance benchmarks are also used to assess the current state of automation technology.

Measures of effectiveness (MOEs) were established for Element 1 and are shown in Table 1. These measures reflect anticipated stakeholder expectations for the concept under study. MOE 1 defines expectations for the vehicle-centric approach that was the subject of study. MOE 2 defines operational scalability expectations. MOE 3 defines expectations for technology maturity. MOE 4 defines expectations for operations within flow corridors. Corridors are an anticipated future requirement for operations in proximity to large regions of restricted airspace, such as an airport that serves commercial transport operations. In this research, flow corridors were defined by restricted SUAs. Maneuverability is expected to be heavily constrained for operations within these volumes. The current reference technology was not designed for operation within corridors, so all research supporting MOE 4 was exploratory.

Table 1. Measures of Effectiveness

Measure	Description
Vehicle-Centric Flight Path Management	
MOE 1A	Aircrews can manage their individual flight paths in UML-4 density airspace while safely sharing the airspace with other aircraft/aircrews that share their flight intent.
MOE 1B	Aircrews can manage their individual flight paths in UML-4 density airspace while safely navigating in the presence of airspace restrictions.
MOE 1C	Aircrews can manage their individual flight paths in UML-4 density airspace while conforming to changing arrival metering constraints.
Operational Scalability	
MOE 2A	Traffic operations are scalable to UML-4 density with no indication of reaching a computational limit within the reference technology.
MOE 2B	Traffic operations are scalable to UML-4 density with no indication of producing a domino effect.
Technology Maturity	
MOE 3	The enabling technology is achievable today. No new foundational breakthroughs are required (excluding breakthroughs required for collaborative and responsible automation).
Operations in Highly Restricted Airspace	
MOE 4	The concept and current technology support operations within traffic corridors not restricted by flow management constraints.

As stated by Reference 15, measures of performance (MOPs) should be used to define the key performance characteristics a system needs to satisfy the associated MOEs when operated in its intended environment. MOPs were used to define the performance benchmarks. Because FPM research is in a requirements exploration phase, FPM MOPs cannot be defined in terms of existing requirements. Instead, MOPs are defined as desired outcomes based on reasonable performance expectations. The quantitative values of the MOPs may serve as starting point for future activities that lead to performance requirements.

The defined MOPs identify function performance in terms of operational behavior at the system level, which requires them to be defined stochastically. A flight test cannot generate enough data samples to satisfy the MOPs. For that reason, positive results from the test can only provide an indication that a MOP will be satisfied with future testing. Results are definitive only if data shows the function under evaluation has failed. Table 2 provides an overview of the eleven MOPs developed for the test. Appendix A provides rationales for the success criteria, analysis approach details, and traceability to the MOEs.

Data metrics were also defined for use in post-test analysis and MOPs evaluation, and are listed in Appendix A.

5.4.2 Element 2: Concept and Technology Advancement

The advancement element provides data to address known unknowns, focusing on development of the representative automation technology and further research of the FPM UML-4 concept. Flight test scenarios were designed to generate data supporting the advancement research. Six foundational research questions were identified for an aircraft-centric implementation of FPM to meet the needs of UML-4 operations. They were used to guide the exploratory analysis of Element 2:

Table 2. Measures of Performance

Measure	Description	Success Criteria
MOP 1	There is a low fraction of false conflict alerts (false positives) and a very low fraction of missed conflict alerts (false negatives).	**i**: Five percent or fewer missed detections **ii**: 20 percent or fewer false detections
MOP 2	Conflicts are detected within the first two CD cycles, measured from when first detection should have occurred.	**i**: True for at least 90% of all occurrences
MOP 3	At least one CR advisory is provided for each detected conflict within the first CR refresh cycle.	**i**: True for at least 90% of all occurrences
MOP 4	Over the design range of CR validity, executed resolutions result in no LOS between the resolving aircraft and any conflicting aircraft.	**i**: Executed resolutions meet or exceed minimum required separation at the CPA between the ownship and all traffic and restricted airspace for at least 90% of detected conflicts. **ii**: 100% of executed resolutions are flyable within the hosting aircraft's performance constraints.
MOP 5	CRs do not result in a repeat conflict between the resolving aircraft and the original conflict aircraft.	**i**: Repeat CDs do not occur for at least 90% of resolved conflicts.
MOP 6	CRs will not create new traffic or airspace conflicts within the design range of CR validity.	**i**: At least 90% of executed CRs create no new conflicts with traffic or airspace regions over the design range of CR validity, excluding conflicts caused by a change of traffic intent within the CR horizon.
MOP 7	CRs will meet any RTA constraint within aircraft performance limits.	**i**: At least 90% of all computed CRs from the first CR cycle will meet an existing RTA within ± 3 seconds if the point of RTA is at least 3 minutes downstream from the hosting aircraft's current position.
MOP 8	FPM technology will prioritize CRs over RTA constraints. If necessary to resolve a traffic conflict, it will provide a CR solution that is out of compliance with the RTA and as close to compliance as possible.	**i**: The conflict is successfully resolved for 90% of all occurrences. **ii**: The RTA constraint will be relaxed by the minimum amount necessary for the hosting aircraft to remain within performance constraints, for 90% of all occurrences.
MOP 9	FPM technology will resolve traffic conflicts in airspace corridors of width between 0.5 and 1.0 NM that are not constrained by RTAs located within the corridor, at expected UML-4 density.	**i**: Conflicts involving corridor operations are resolvable for at least 90% of detected conflicts. **ii**: Executed resolutions meet or exceed minimum required separation at the CPA between the ownship and all traffic and restricted airspace for at least 90% of detected conflicts.
MOP 10	FPM technology will complete all CR computations and display CR alerts within the first 15 seconds of each refresh cycle at UML-4 traffic density.	**i**: Complete CR computations and display resolution advisories within 15 seconds for 100% of all occurrences.
MOP 11	FPM technology will provide operators with sufficient time to evaluate and execute a resolution before a new set of resolution advisories become available to evaluate.	**i**: For maneuvers using baseline time parameters, resolution advisories are displayed to the operator for execution for at least 15 seconds before being removed or replaced, for 90% of all cases.

1. What CR success rate should FPM be required to achieve?

2. What separation standards will ensure adequate traffic flow for expected UML-4 traffic density?

3. At UML-4 traffic densities, will human operator attention and reaction time (and their variability) be adequate to allow human

4. What is the minimum traffic separation (made up of a separation standard plus any necessary buffers) required by FPM to achieve acceptable CD and CR performance?

5. Can FPM technology (automation plus associated communications) adequately avoid triggering a domino effect? A CR that spawns new conflicts and causes a chain reaction is an example of domino effect.

6. How do FPM functions designed for unconstrained airspace perform in constrained airspace such as UAM corridors?

Although a complete resolution of these foundational questions is far beyond the scope of the current research, the advancement research element has the goal of providing initial insights toward their resolution. The recorded data and pilot responses to surveys and interviews were analyzed to identify initial trends and develop initial insights. The following topics were subjects of analysis.

- **Trajectory Conformance** - Analysis of AOP's trajectory generation that impacted the ownship's trajectory conformance.

- **Conflict Detection** - Analysis of CD as a function of other parameters, as permitted by test data availability.

- **Conflict Resolution** - Analysis of CR as a function of other parameters, as permitted by test data availability.

- **Conflict Resolution Compute Time** - Analysis of the computation time required for PBGA resolutions as a function of traffic density, time-to-go, phase of flight, and other potential factors.

- **Arrival Time Compliance** - Analysis of AOP's capability to meet RTAs, for both CRs and in response to RTA changes. Parametric analysis may include RTA compliance as a function of RTA change magnitude and time to the RTA point.

- **Performance in Corridors** - Analysis of AOP's capability to perform its CD, CR, CP, and RTA compliance functions while constrained by airspace corridors.

- **Pilot Assessment** - Analysis of research pilot feedback related to quality and characteristics of CR solutions. Analysis was based on end-of-run survey data and end-of-test pilot interviews.

The above analyses were used to gain insights and draw initial conclusions regarding the foundational research questions.

5.4.3 Element 3: Discovery

Because emergent behavior is expected, unexpected technology performance and system behaviors may be revealed when tested in a relevant environment. The Discovery research element attempts to uncover some of these unknowns by providing stressor scenarios. Stressor scenarios were designed to exceed expected system performance and reveal unexpected behaviors, trends, or failure modes. Stressor scenarios included automation settings modified for long and short lookahead horizons, short (pop-up) traffic conflicts, traffic rules violations (e.g., traffic changes its plan and thereby creates a conflict), and high-density traffic.

5.4.4 Element 4: Simulation Validation

The simulation validation element provides flight test data to support validation and refinement of AAM simulations and simulation models. Simulation validations seek to increase confidence in simulation results that scale operations to levels that cannot be performed in flight due to large scale and high cost. Flight test data were used to perform a cursory simulation validation. In post-flight simulations, all relevant environmental conditions experienced during flight data collection were used to duplicate flight test runs. Stressor scenarios and unexpected results cases received special attention. An analysis of results determined a level of confidence in the current simulation and provided recommendations for future simulation validation research.

6 Flight Test System

The objective of the FPM portion of the flight test was to evaluate FPM automation in the context of the four research elements defined in Section 5.4. Due to the expensive and complex nature of flight testing, significant effort was put towards designing the flight test system to ensure that the intended research could be conducted successfully. This section will first cover the test requirements used to define the flight test system needed to conduct FPM research. Next, the design of an LVC environment for UML-4 operations will be described, along with the test system developed to execute the research flights. Lastly, the method for creating intentional traffic conflicts between the research aircraft will be summarized. Sections 6.2, 6.3, and 6.4 were originally published in Reference 18 and have been included with additional information.

6.1 FPM Test Requirements

The following test requirements were developed to guide the design and development of the flight test system and environment:

1. Technology shall be tested in flight onboard at least one host “ownship” aircraft.

2. Technology shall be installed on a research computing platform. An avionics platform installation is not required for TRL-5.

3. Technology shall have a dedicated pilot UI for functional evaluation purposes (referred to as an engineering UI) on a temporary tablet computer. Forward field of view integration and a certifiable UI design are not required for TRL-5.

4. Technology shall receive the ownship data from the host aircraft avionics.

5. Technology shall receive traffic data through external data transmissions via vehicle-to-vehicle and/or ground-to-vehicle links.

6. External data transmission links shall use either actual systems or proxy systems. Certified links with the needed functionality may not exist and are therefore not required for TRL-5.

7. Traffic seen by the technology shall include at least one actual "intruder" aircraft. The rest of the "background traffic" may be simulated. Hundreds of live aircraft are not required for TRL-5.

8. The quantity of background traffic shall be representative of UML-4.

9. Background traffic shall fly routes representative of expected UML-4 operations in a metro environment. Live playback of recorded tracks of the background traffic are acceptable for TRL-5.

10. Live state and trajectory intent data shall be sourced from the intruder and reflect its actual state and commanded intent. If intruder changes intent, the intent message shall change to match the new intent.

11. 4D navigation performance of the ownship and intruder shall be established in advance of the flight test.

12. Dynamic constraints (e.g., RTAs, restricted airspace status, weather) shall be transmitted from an external source to the ownship and routed to the technology. Data sources may be simulated for TRL-5.

13. Wind field data as needed by the technology shall be sourced from an established data provider for the actual flight day and may be preloaded in the technology prior to flight. Live access to weather data is not required for TRL-5.

6.2 Flight Test Environment

Early in the planning, two decisions helped to create the desired future operational environment of UML-4: choosing flight test aircraft and defining an LVC airspace operating environment. These decisions were used to develop a concept of test operations. This concept of test operations was specifically designed to define the high-level FPM use case for the flight test, which provided a baseline for functional, performance, operational, and environmental requirements for data collection. This sub section describes how the flight test environment was designed to conduct relevant FPM research for a future operational environment while flying within the current operational environment.

6.2.1 Flight Test Aircraft

To test the FPM technology, two research aircraft provided by Sikorsky Aircraft Company were utilized in the flight test: the SARA S-76B helicopter and the Optionally Piloted Vehicle (OPV) S-70 Black Hawk, both pictured in Figure 2. All flight operations involving SARA and OPV were piloted for safety as required to represent the future UML-4 concept. In addition to the pilot in command, each aircraft's crew included a NASA research pilot onboard who operated the FPM and IAS systems.

These vehicles served as surrogates for future eVTOL aircraft. Both aircraft had operational limits when operating with their research systems active. The maximum calibrated airspeed (CAS) for both SARA and OPV was 120 kt, while the max ground speed for both aircraft was 145 kt. Additionally, both aircraft operated under a 50 kt CAS minimum limit. These limits had to be considered in the design of the flight test. For this

(a) SARA aircraft. **(b)** OPV aircraft.

Figure 2. SARA and OPV aircraft.

flight test, SARA was referred to as the ownship aircraft hosting AOP and the OPV aircraft was referred to as the intruder aircraft which had the role of creating the planned encounters with the ownship.

6.2.2 Airspace Operating Environment

The two flight test aircraft operated in a mixed-reality, medium-traffic-density airspace with a minimum of 225 simulated traffic aircraft flying routes defined by the simulated UML-4 operating environment of the FPM-1 and FPM-2 activities (Ref. 17). This LVC approach was identified as a safe and cost-effective way to achieve the required traffic density levels to represent UML-4. Figure 3 illustrates the simulated UML-4 operating environment used for the flight test. The environment was based on studies conducted by NASA Ames Research Center and the Virginia Tech Air Transportation Systems Laboratory (Ref. 19). The region of airspace surrounding Dallas Fort Worth (DFW) was selected as the more challenging of the two cases documented in the reference. Unusable airspace, represented as "area hazards" to avoid, is significant, making up 11.6% of total airspace for the modeled DFW region. These restrictions, shown as yellow regions in Figure 3, represent airspace heavily used by commercial traffic through DFW and other area airports.

Future demand was modeled, and vertiports were identified and located based on mapping the anticipated demand to the region. Routes between vertiports were designed to minimize path distance while avoiding restricted airspace, but were not restrictive (i.e., UAM flights were free to navigate in the open airspace). Corridors were added to allow for less circuitous paths between vertiports. This simulated operating environment was used for previous FPM batch studies and simulations to represent a UML-4 environment (Refs. 16, 17).

All aircraft, including the flight test aircraft (SARA and OPV) and simulated background traffic, exchanged aircraft state information (current position, track, groundspeed, altitude, and vertical speed) and flight path intent information (full 4D trajectory to their destination). To facilitate communication of this information between the two flight test aircraft, SARA and the OPV aircraft were equipped with Automatic Dependent Surveillance–Broadcast (ADS-B) Out/In and a proprietary Sikorsky telemetry system. ADS-B was used to share state information between the two flight test aircraft, and the Sikorsky telemetry system was used as a proxy for future communication technologies to share intent and other information.

6.2.3 Test Area and Constraints

The test area was established over the Long Island Sound with flights originating from Sikorsky Memorial Airport (KBDR). NASA and Sikorsky partnered to define a test area to contain all flight-test operations, as shown by the magenta box in Figure 4. The rectangular area was centered immediately south of KBDR with

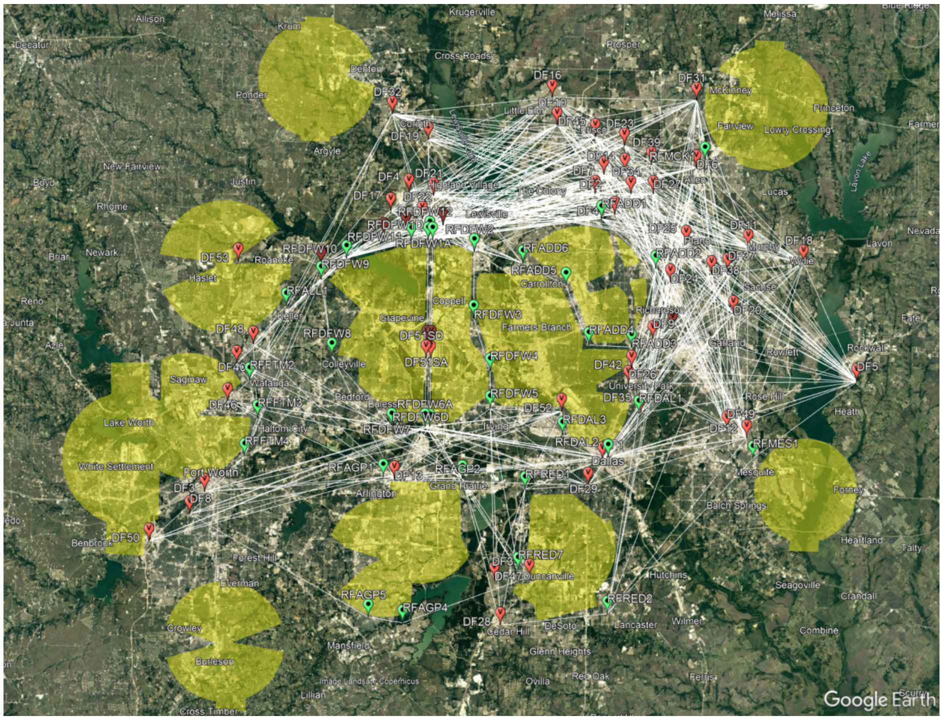


Figure 3. Virtual UAM UML-4 Operating Environment

a length of approximately 27.5 statute miles running southwest to northeast along the sound and a width of approximately 13 statute miles. SARA and OPV aircraft remained within 20 NM of each other and within 25 NM of KBDR to ensure a strong data link and to minimize interaction with nearby commercial operations. Within the 25 NM radius of KBDR (red circle) were New York Class B airspace (dark blue), Long Island MacArthur Airport Class C airspace (purple), and the Bridgeport/Sikorsky Airport Class D airspace (light blue), along with many others as seen in Figure 4.

Overlapping the test area with Class D KBDR airspace was determined to be acceptable due to the familiarity of air traffic controllers at KBDR with Sikorsky flight test operations. The flight crews of the SARA and OPV aircraft coordinated with the KBDR tower as necessary, though Class D airspace was avoided whenever possible to simplify the pilot's workload. Additionally, New Haven Class D airspace intersected the northeast section of the Sikorsky boundary box and was avoided completely to minimize air traffic control coordination.

6.2.4 Final Test Environment Design and Mission Rules

To incorporate the simulated UML-4 DFW operating environment in the flight test, the environment was transposed and rotated to a geographical coordinate near KBDR. A particularly aircraft-dense section of the environment was chosen to overlay within the Sikorsky boundary to ensure UML-4 traffic density was represented. This section along with vertiports (yellow pushpins), vertiport routes (white lines), and restricted airspace regions (yellow polygons) can be found in Figure 5. The rotation angle varied between test points based on the ownship trajectory orientation and the purpose of the encounter. For some test groups, the restricted airspace needed to be repositioned to create certain constraints or encounter geometries. One such maneuver required one of the corridors from the UML-4 operating environment to be positioned within the magenta boundary, which led to a new transposition of the operating environment.

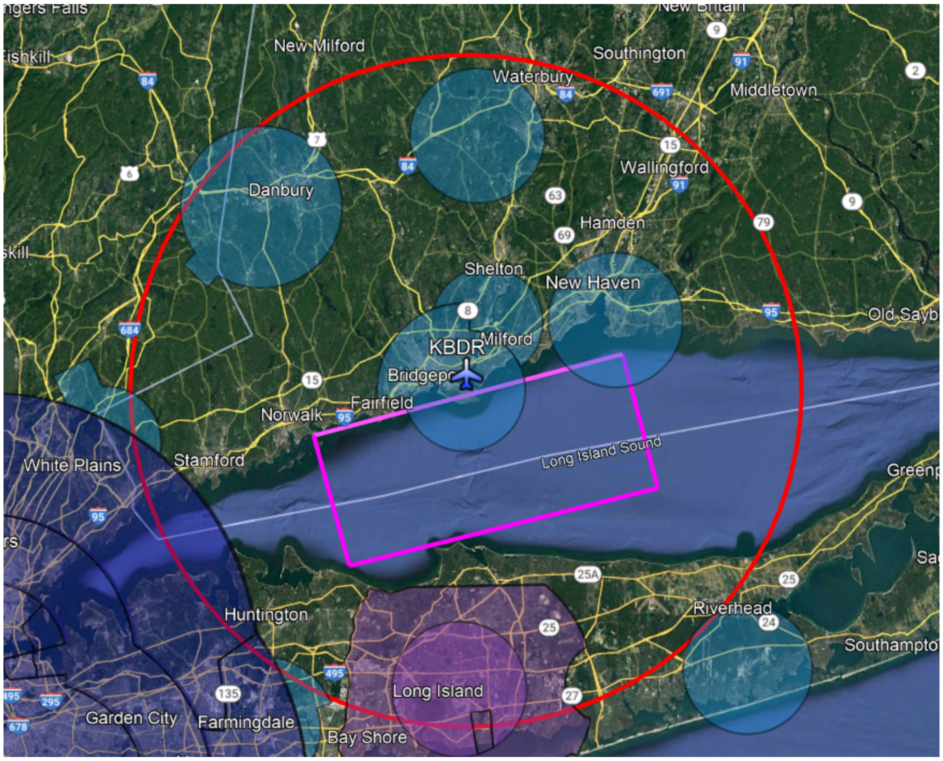


Figure 4. Airspace Surrounding Sikorsky Memorial Airport (KBDR)

Rotating the DFW environment not only repositioned the restricted airspace regions but also changed the cardinal direction of the virtual traffic. Due to how the simulated background traffic is currently generated, changing the cardinal direction of the traffic at its origin point also altered at what altitude the traffic flew. All simulated background traffic flew at either 1500 ft, 2000 ft, or 2500 ft if the aircraft was traveling $0°$-$120°$, $120°$-$240°$, or $240°$-$360°$, respectively. This was accounted for when designing each maneuver to ensure the ownship would be influenced by an adequate amount of co-altitude simulated background traffic.

For safety, flight mission rules were developed by the flight test operations team at NASA Armstrong Flight Research Center. Specific to FPM, all FPM encounter geometries were designed so the CPA of the SARA flight plan and the OPV flight plan was never less than 0.1 NM laterally. There was no equivalent vertical offset. Additionally, encounters between the SARA and OPV aircraft were designed so separation would never actually be lost; maneuvers were procedurally terminated well prior to separation loss (at least 30 seconds prior to LOS). If the two aircraft were separated by less than 500 ft vertically, visual acquisition of the other aircraft by one of the pilots was required within 0.75 NM. The two aircraft were only allowed to be within separation standards if a AOP generated resolution had been executed and the two aircraft were in the process of diverging from each other. Unless otherwise specified, maneuvers were designed to be flown at a 2000 ft altitude. However, all maneuvers were subject to a 500 ft above ground level altitude floor. Furthermore, if any anomalous behavior from the research maneuver or test system was experienced, that run was subject to termination. Finally, if any non-participating aircraft were observed within 2000 ft vertically and 3 NM laterally of either flight test aircraft or the non-participating aircraft and could not be determined to be a non-factor (i.e., flying a non-convergent flight path with either test aircraft), the test run was terminated. All present personnel had the ability to terminate a run that they deemed unsafe at any time using the NASA “Knock-It-Off” best practice (Ref. 20).

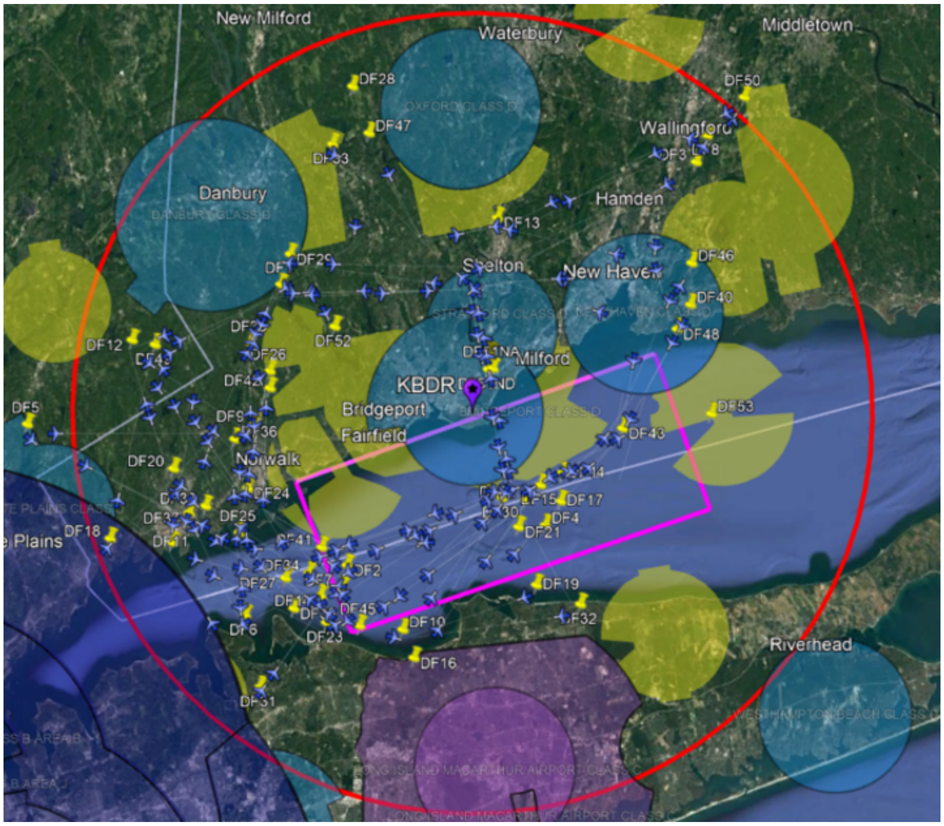


Figure 5. Dallas/Fort Worth (DFW) UML-4 Database Overlaid on the Surrounding KBDR Airspace

6.3 Test System Architecture and Summary

IAS-1 was a multi-center, multi-organization effort that required collaboration between several research teams and technical personnel, as well as integration of multiple sophisticated technologies. Significant integration efforts were needed to ensure that these complex technologies were harmoniously integrated to produce the functional test system required to successfully meet the planned research goals. The development and integration of the flight test system occurred over the course of a year through a series of preparation flights. These initial flights allowed the NASA teams and Sikorsky to test system integration, identify and resolve technical issues, and solidify the operational methodology. As a result, the finalized flight test system was developed. More details on the early flight test campaigns can be found in Reference 21. The test system constituted three main elements represented by the predominant boxes in Figure 6: the ownship aircraft (SARA), the intruder aircraft (OPV), and the Ground Control Station (GCS). Each element comprised of a set of research systems and technologies, which were interconnected by various data or communication links. The arrows in Figure 6 show the interaction type between the various technologies and human operators.

6.3.1 AOP Engine

The AOP software generated the ownship 4D trajectory intent, detected conflicts with traffic and airspace hazards along the active route, and computed trajectory change solutions (lateral, vertical, speed, or hybrid) in response to detected conflicts. Additionally, AOP generated AOP health monitoring data for situation awareness on current data reliability. AOP also logged a collection of data files for use in post-test research analysis activities.

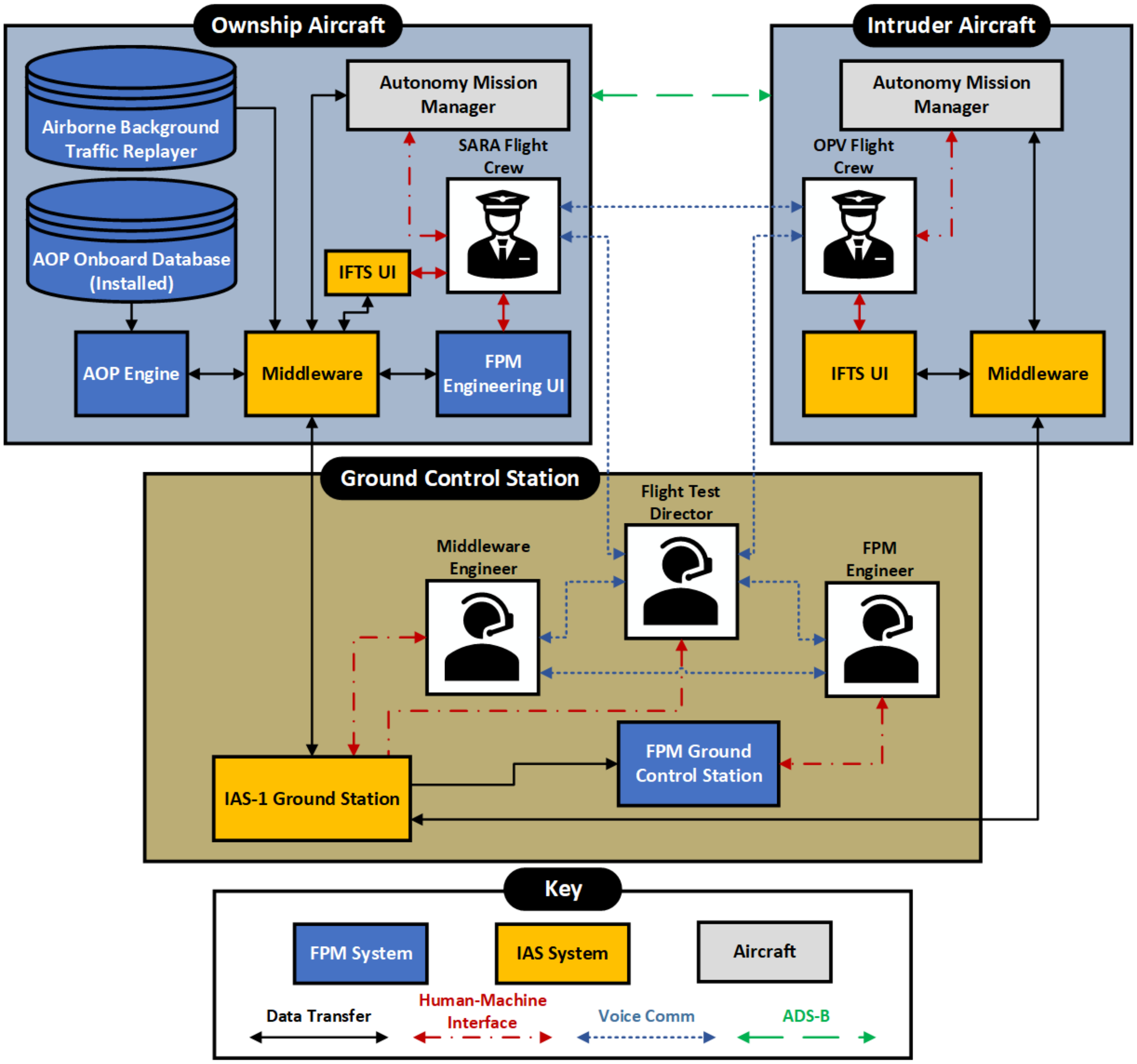
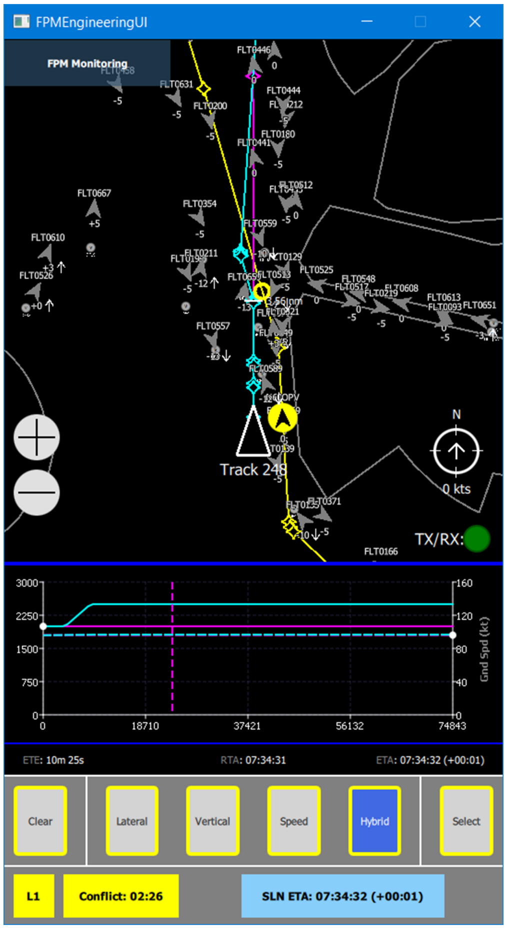


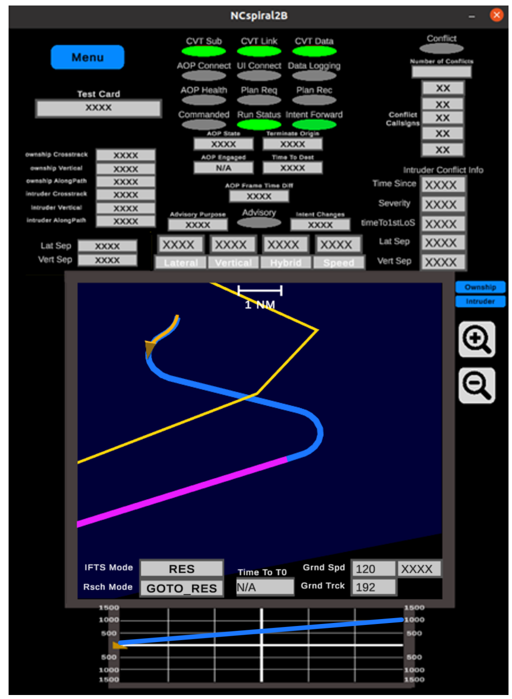
Figure 6. Simplified Flight Test System Overview

6.3.2 FPM Engineering UI

Flight test research pilots were provided a UI to the FPM automation to allow them to evaluate the practicality of the technology's solutions. Use of the automation technology to support human decision-making and the evaluation of procedures for use were outside the scope of the flight test, so the interface was not designed to support such research. Referred to as the FPM Engineering UI (Figure 7a), the interface was provided to the research pilot on a touchscreen tablet computer that communicated with the AOP software engine residing on a separate computer. It was responsible for providing a graphical display of the current operational situation and environment, and for providing a mechanism to preview, select, and execute AOP-generated trajectory change solutions. The interface provides a plan view and profile view of the environment in proximity to the hosting aircraft. The plan view displays real and virtual traffic participating in the test. It also shows airspace regions to avoid, which are represented by polygons. Additional information is displayed



(a) FPM Engineering UI showing an acute crosser encounter.



(b) The FPM Ground Control Station UI.

Figure 7. FPM Engineering UI and the FPM Ground Control Station

as it is relevant to the situation, such as an RTA constraint. The example shown in Figure 7a displays an acute crossing conflict with the intruder, with a predicted LOS location shown by the yellow circle near the center of the plan view. The intruder aircraft is highlighted because it is in conflict with the hosting aircraft. Magenta lines represent the hosting aircraft's active trajectory, and cyan lines display one of the CR advisories provided by AOP.

6.3.3 FPM Ground Control Station

Like the FPM Engineering UI, the FPM GCS provided a graphical display of the LVC operational situation from the ownship's perspective (e.g., present position, current route, winds, conflicts, applicable constraints). It also provided a mechanism to view (no selection functionality) AOP-generated trajectory solutions via graphical interface, as well as displayed AOP health monitoring data. The display was only used by the FPM engineer on the ground, not the ownship or intruder pilot. The FPM GCS graphical display is shown in Figure 7b, which depicts the target ownship maneuver trajectory (magenta track), automated guidance to the maneuver start point (blue track), and the track actually flown by the ownship (yellow track behind the ownship chevron). The bright yellow outline that the blue track passes through is the boundary of an SUA from the simulated DFW operating environment. SUAs within the virtual environment were disregarded as obstacles through all flight operations prior to AOP engaging at the maneuver start point.

6.3.4 AOP Onboard Database

The AOP Onboard Database was a collection of AOP data specific to the FPM research test groups. The database included AOP configuration files, flight maneuvers and scenario information, wind data, and avoidance polygons. This database only outputted to AOP via disk.

6.3.5 Background Traffic Replayer

The airborne Background Traffic Replayer provided the virtual traffic playback needed to achieve UML-4 traffic density to conduct FPM research. It provided playback background traffic state data at 1 Hz with intent data, which was distributed only when the background traffic intent was created or changed. The Background Traffic Replayer provided state playback in the air to the SARA aircraft. This was done to reduce communication flight test bandwidth requirements.

6.3.6 Middleware

AOP was integrated for flight test by the IAS team utilizing the Middleware. The Middleware provided an aircraft-agnostic method to integrate automation technologies, in which the aircraft specific interfaces were behind an abstraction layer. This abstraction layer allowed the automation technology to be integrated with the Middleware, and then become insulated to changes to the sources of the ownship and the intruder states as well as changes to the way that the autopilot was commanded. The automation technologies were integrated within a “monitor” plugin that was hosted by the Middleware. A monitor is the mechanism in the Middleware architecture to host systems that generate and send commands based on a run time assurance framework as described in Reference 22. The Middleware was able to host multiple monitors while keeping the behavior predictable for individual monitors. The Middleware was utilized on both SARA acting as the ownship and OPV acting as the intruder. The IAS Flight Test System was also included in the Middleware as a monitor. This system allowed the AOP test scenarios to be configured and executed from the Middleware Engineer UI (Figure 8) by a Middleware Engineer in such a way that both the ownship and the intruder were at the desired location and a coordinated time. This allowed for precise test engagements to be coordinated that allowed the FPM team to repeatably execute test encounters.

The AOP monitor that was hosted by the Middleware running on an Intel NUC (Next Unit of Computing) platform acted as a central hub for the AOP system (Figure 9), which consisted of AOP hosted on a second Intel NUC, the FPM Engineering UI which was hosted on a Getac F110 tablet, and an FPM GCS (Figure 7). The AOP monitor was responsible for providing the AOP system with the ownship state, the intruder state, the intruder intent, virtual background traffic state and virtual background traffic intent. The AOP monitor set up each test encounter based on predefined scenarios which involved configuring virtual background traffic state and intent playback as well as configuring the AOP engine for that specific test maneuver. The intruder intent was received from the Middleware instance on OPV via a message repeater in the GCS. The AOP monitor wrote key parameters to a Middleware-provided database called the Current Value Table that was written to a log and was also streamed to the ground to feed data to the FPM GCS. The Middleware handled the conversion of the time-based trajectories that AOP generated in the Efficient Universal Trajectory Language (EUTL) (Ref. 14) format to the velocity-based commands that the SARA and OPV Autonomy Mission Manager (AMM) expected. This was done through a velocity controller that minimized time error

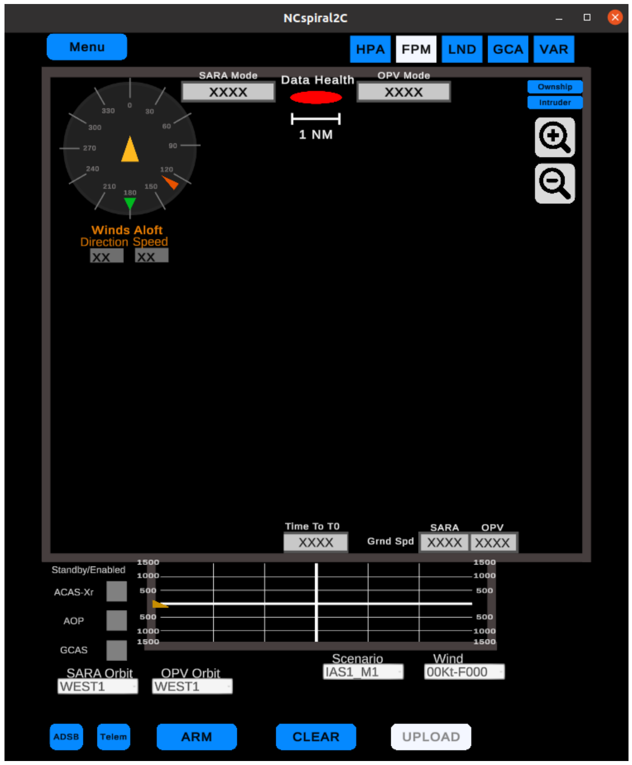


Figure 8. Middleware Engineer UI

via velocity modulation. These velocity commands are then communicated by the AMM to the Vehicle Mission Computer, which is responsible for executing them.

6.3.7 IAS-1 Ground Station

The IAS-1 Ground Station, which contained a critical Middleware Engineer UI system, was an interactive multi-window interface that allowed the IAS test engineers to transmit test group specific scenario and AOP configuration data to the IAS Middleware, which acted as a centralized data hub, for distribution (Figure 8). The Middlware Engineer UI system specifically played a role in setting up coordinated guidance to the start of each FPM maneuver. It was also responsible for receiving AOP-generated and selected trajectory change requests that were sent to the FPM GCS.

6.3.8 Autonomy Mission Manager

Both Sikorsky research aircraft were fitted with the AMM system, which is what allowed the aircraft to automatically execute a 4D trajectory. The AMM was developed by Sikorsky to be their research autonomy system and is hosted on the onboard High Performance Computer, along with the other software onboard the aircraft. To fly AOP-generated 4D trajectories with the Sikorsky research aircraft, the Middleware interfaced with the AMM system, converting information into a format readable by the AAM. AMM interfaces with the aircraft systems by communicating with the Vehicle Management Computer, which hosts safety-critical flight software.

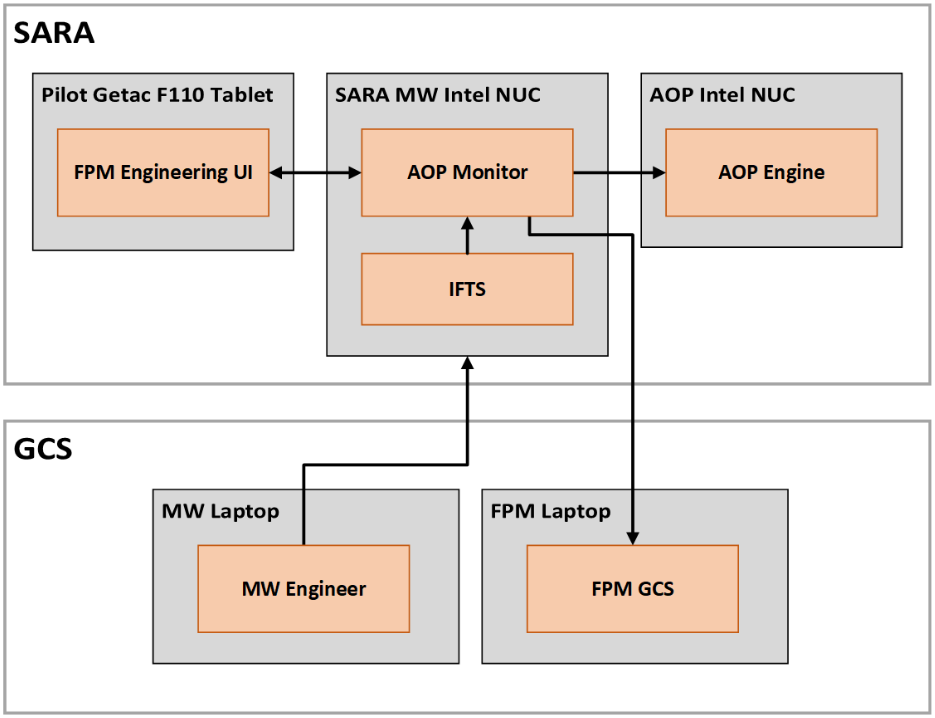


Figure 9. Middleware Monitor Configuration

6.4 Creating Intentional Traffic Conflicts

One of the most difficult challenges to be solved for the flight test was reliably creating conflicts between the ownship and the intruder so AOP could resolve them. As explained below, specifying a predetermined 4D trajectory for each aircraft to follow was found to require extra steps to create a conflict that meets test objectives. The need for preprocessing to create the conflict also resulted in a requirement for the two aircraft to coordinate their starting locations and speeds, as well as a procedure for arriving at their starting points simultaneously.

6.4.1 Encounter Scenario Generation

In order to construct 4D trajectories that ensured the intended encounter between the two flight test aircraft, a scenario generation tool that was used to generate the FPM-2 random scenarios was extended to also produce 4D trajectories for the ownship and intruder for the specific encounter geometry (crossing, head-on, overtaking, and overtaken) and other factors (RTA changes and additional, hand-placed constraining traffic) required for each cell in the test matrix. The scenario tool is built around the same BBTG that is used in AOP (Refs. 10, 23). The BBTG is used to compute kinematically complete, flyable trajectories for each UAM vehicle type and origin-destination pair needed in the scenario. For background traffic, these trajectories were used to schedule non-conflicting operations at each vertiport. For the ownship and intruder, these trajectories were used to align the flights to each other in time at the conflict point and then extract the necessary initial conditions (position and velocity) to stage the vehicles a specified time before the point of interest. For background aircraft and the intruder, the time-aligned 4D trajectories were recorded using the EUTL for use by the Airborne Background Traffic Replayer and the Middleware running on the intruder (Ref. 14). For the ownship, a mission specification was captured and passed to AOP at the beginning of the scenario to enable AOP to recreate the desired ownship trajectory using AOP's copy of the BBTG. To ensure that the ownship did not encounter a conflict with a virtual background traffic before the intended encounter with the Intruder aircraft, the scenario tool checked all background traffic trajectories for conflict with the ownship trajectory beyond the initial condition and removed the few conflicting background flights.

Because the generated scenarios are based on ground-referenced 4D trajectories that dictate what ground speed each vehicle must fly, the scenario generation process must include a wind prediction so that the vehicle trajectories reflect the desired vehicle airspeeds. The scenario data needed to be generated weeks before the flight test, so scenario data for each test matrix cell was generated using a set of wind condition permutations to create a database of flyable scenario data indexed by wind condition. Because the 4D guidance systems onboard the two test vehicles would be adjusting airspeed to conform to the ground speeds prescribed by the 4D trajectories, it was only necessary to use a wind condition matrix fine enough to be within about 10 kt of the actual winds present during the sortie. The completed database resulted in the ability to select a maneuver configuration based on the wind conditions during that sortie to successfully achieve the desired conflict. The full list of wind speed and direction combinations generated for each maneuver totaled 37, including a 0 kt case (Table 3).

Table 3. Wind speed and direction combinations for each maneuver.

Wind Speed (kt)	Wind Direction (degrees)
0	N/A
10	000, 060, 120, 180, 240, 300
20	000, 060, 120, 180, 240, 300
30	000, 030, 060, 090, 120, 150, 180, 210, 240, 270, 300, 330
40	000, 030, 060, 090, 120, 150, 180, 210, 240, 270, 300, 330

6.4.2 Establishing Starting Points

The scenario generation tool helped ensure that, despite the airspeed and ground speed limits of the Sikorsky research aircraft, each maneuver could almost always be flown if the test day winds were included within the matrix of compensated winds (Table 3). In a few specific wind conditions and maneuver combinations, it was not possible to fly the maneuver without exceeding the Sikorsky ground speed limits. These specific instances were noted before the test and precautions were incorporated into the execution plan for the test. The tool used the intended location of separation loss as an input, and it computed the trajectories of each aircraft that would create the loss. It therefore determined the starting locations, which were a function of wind speed and direction at the time of the test. These starting locations were referred to as the T_0 points.

A T_0 point was generated for both aircraft for a given maneuver and consisted of a latitude, longitude, altitude, ground speed, and ground track angle where the aircraft was required to be at the start of the maneuver to ensure the intended scenario was achieved. Depending on the effect of the wind compensation, the T_0 point would shift approximately one to two nautical miles in either direction along the initial track angle at the start of the maneuver trajectory. The combination of location and speed at T_0 allowed the two aircraft to arrive at the intended location in time to coordinate the desired CPA based on the specific maneuver's objective.

Trajectories computed by different trajectory generators will likely not be identical, even if they are provided the same trajectory constraints. Different trajectory generation algorithms use differing approaches and assumptions in their computations. In preparation for the flight test, it was discovered that specifying a

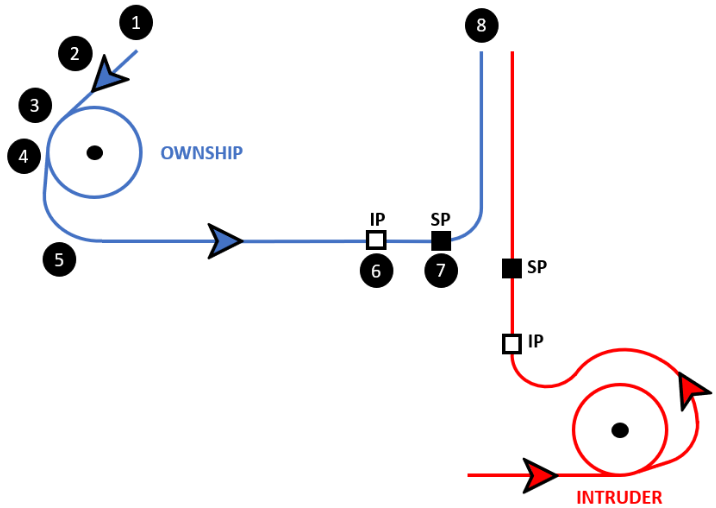


Figure 10. Operational Flow of IFTS Navigation to Run Start Points (SP)

trajectory that was not computed by AOP's trajectory generator, BBTG, often resulted in an unintentional removal of the conflict as the trajectory was recomputed by AOP. In computing the trajectories of the aircraft, the scenario generation tool therefore executed the BBTG to ensure no trivial resolutions removed the conflict. Additional details may be found in Reference 10, pp. 8–9.

6.4.3 Navigating to the Starting Points On-Time

The flight test setup and automation is explained in detail in Reference 21, but here we provide a brief overview of the methods. In addition to the challenge of creating the correct T_0 points in conjunction with the desired 4D trajectory to successfully create the intended conflict, functionality was needed to ensure the two aircraft arrived at T_0 simultaneously to start the test run. Integrated Flight Test System (IFTS) was responsible for configuring the test runs for AOP and ensuring that SARA and OPV were at the desired location, and velocity vector at a known time for T_0. T_0 was a foundational parameter for each AOP test scenario, that could only be computed and coordinated with both SARA and OPV moments before the scenario began. To accomplish this coordination, an operational flow (Figure 10) was created that allowed both aircraft to loiter at a known and predefined orbit location while the coordination was completed prior to release from the orbits.

The operation flow used to enable the execution of FPM maneuvers was completed with the following steps:

1. Idle: System idle and waiting for test scenario setup. During this stage, the aircraft will fly straight and level allowing for an accurate wind estimate.

2. Orbit Staged: The test orbit location for both the ownship and the intruder is selected and staged. The specific test maneuver is selected, and the wind estimate is used to select between the quantified set of wind speeds and directions.

3. Orbit Approach: The test aircraft is approaching to enter the desired orbit location tangentially at the desired orbit direction.

4. Orbit Idle: Both aircraft loiter in the desired orbit locations until both are ready to begin the test run.

5. Research Approach Coordination: IFTS on both aircraft computes the time it would take to arrive at the Initial Point (IP) when flying straight and level, which is determined to be 10 seconds away from the starting point. The starting point corresponds to the same location at T_0. The aircraft with the longest time to arrive at the IP is used to set the T_0 time, and that aircraft is staged and ready to proceed with its computed route pending coordination with the other aircraft. The aircraft with the lesser time iterates through one second intervals in the orbit, and then recomputes the time to arrive at the IP until it is near the first aircraft's longer period to T_0.

6. Research Approach: Both aircraft finish coordinating a T_0 time, and the approach trajectories are commanded.

7. Research Mode: T_0 is achieved, so the Middleware hands control over to the AOP monitor to fly the FPM maneuver.

8. End Mode: FPM run is completed and the test run ends.

7 Flight Test Design and Execution

A set of eight test groups were developed to conduct FPM research during the flight test, each of which was made up of several test maneuvers. The maneuvers were designed to account for the limitations of the test area and research aircraft, and incorporated the insights made from the outcomes of preliminary flight testing. Each test group was crafted to gather flight test data to support four research elements (see Section 5.4) to help increase the understanding of AOP's functionality and its effectiveness in a relevant environment. The FPM research portion of the flight test was executed with the intention of collecting a comprehensive set of data to analyze.

The following section expands upon the summary of the flight test design provided in Reference 18. Sections 7.1 and 7.2 were originally published in Reference 18 and are included with additional information. The four research elements were supported by two types of test scenarios: nominal operations and stressor operations. Each of the eight test groups were designed for a primary focus on one or the other, although a shared focus was possible. Nominal test scenarios contributed to the Function Verification research element. Maneuvers were designed to stay within the range of operational conditions and constraints for which the technology under test was designed. The nominal scenarios were used to evaluate algorithm performance and system performance in meeting deconfliction and time-compliance objectives. Stressor scenarios were used to explore concept behaviors and limits of the enabling technology. With a goal to gain insight from failure, stressor maneuvers were designed with the expectation that some would fail. Technology failures would contribute to development (Concept and Technology Advancement and Simulation Validation research elements) and concept failures or unexpected behaviors would contribute to the Discovery research element as well as Simulation Validation.

7.1 Test Groups and Maneuvers

To successfully meet specific research purposes, each test group consisted of a set of maneuvers with varying encounter geometries, altitude changes, AOP parameters, and execution instructions. Most maneuvers were designed to create encounters between the two aircraft that caused a conflict for AOP to resolve. The test groups were prioritized as shown in numerical order below to ensure that high-priority tests were completed within the flight hour constraints. With 52 unique test points available to fly, there were not enough available flight hours to fully execute all eight test groups. The test day run priority was adjusted during the first half of the flight test campaign to ensure a sufficient sampling of each test group was collected before revisiting any unfinished test groups.

- Test Group 1 - Conflict Detection: Designed to explore AOP's CD functionality by flying encounters that intentionally create a conflict with an intruder aircraft. The purpose of Test Group 1 was to verify the CD function performs as expected. It also had a goal to generate large quantities of CR data given limited flight test hours. This was achieved by allowing AOP to continuously generate resolution advisories as the two aircraft converged. Pilots were instructed not to execute any advisories. Because they were not executed, resolution advisories generated by Test Group 1 could not be verified to resolve conflicts, and therefore were not used for that part of CR function verification. The test aircraft were allowed to converge until mission safety rules required an end to the maneuver. By executing the maneuver for the maximum amount of time before LOS, a stressor objective was also achieved: data were collected to determine how close to the point of first separation loss AOP could generate CR solutions. Test Group 1 was also used to provide a final verification that both test aircraft flew in conformance with their 4D trajectory guidance. Finally, as the first test group flown, Group 1 provided a final verification of test procedures. Test Group 1 was made up of nine unique maneuvers.

- Test Group 2 - Conflict Resolution and Prevention: Sought to verify CR and CP functions performance as expected. Test Group 2 used the same nine maneuvers from Test Group 1, but pilots were instructed to select a resolution trajectory from the set of resolution advisories provided by AOP. Test runs were typically continued until the test aircraft were beyond their CPA and trajectories were clearly diverging. Because a trajectory was executed, the data recorded was limited to a single trajectory, but that resolution was important in verifying that it resolved the conflict without creating new conflicts. Pilot insights and preferences were collected regarding their decision making in selecting a resolution from the set offered by AOP. As confidence was gained in test aircraft trajectory conformance, flight test time was conserved by reducing the requirement for all test runs to continue the test maneuver until the aircraft had diverged.

- Test Group 3 - RTA Change Compliance: Tested AOP's ability to detect and resolve an RTA change while staying deconflicted. The purpose of Test Group 3 was to verify that AOP can detect non-conformance of active trajectory with an RTA that has been changed, as well as compute a new trajectory that complies with the changed RTA while meeting airspace constraints and without generating traffic conflicts. As in Group 1, Test Group 3 runs did not execute AOP resolutions so each run would generate large amounts of resolution data. Also, similar to Test Group 1, this group provided data for a stressor objective to determine how close to the RTA point AOP is capable of generating resolution trajectories. During the test, Group 3 occasionally approached boundaries of the test range because of the longer time duration of the runs. Test Group 3 was made up of five unique maneuvers.

- Test Group 4 - RTA Change Compliance with Traffic Conflict: Evaluated AOP's ability to detect and resolve an RTA change that creates a conflict with traffic. Maneuvers were designed with the intruder on the same arrival path as the ownship, in conflicts when an RTA change was received. Test Group 4 involved selection and execution of a AOP resolution advisory for function verification purposes and to collect pilot opinions, as was done for Group 2. One test maneuver was designed specifically to test AOP's intended function to relax the RTA constraint if achieving both CR and RTA compliance is not possible. Test Group 4 was made up of three unique maneuvers.

- Test Group 5 - Time Parameters Variation: Investigated the impact of changing internal AOP time horizon parameters on detecting, resolving, and preventing traffic conflicts in a UML-4 operating environment. The purpose of Test Group 5 was to gather in-flight data that would provide insight for later optimization of the AOP parameters. It also had the goal to provide a reasonably comprehensive data set for post-test validation of FPM simulations. Test Group 5 was designed to have twelve unique maneuvers, which when combined with Group 2's runs that contained default parameter settings, would produce a sixteen-maneuver data set. Unfortunately, not enough flight test time was available to complete all planned maneuvers. Five maneuvers were completed for Group 5.

  Four primary time parameters were varied across Group 5 maneuvers: CD lookahead time (time horizon for detection), CR buffer (lookahead time beyond the CD lookahead time horizon, where CR horizon is the sum of both), freeze horizon for CR (the period for which a new flight plan must agree with the current flight plan), and refresh interval (the frequency of computing CR). In addition to these parameters, three alert level time configurations were varied, which represent different degrees of urgency regarding the current conflict. Sets of values of these parameters were classified as "baseline," "short," "long," or "very long" as shown in Table 4.

Table 4. Time Parameters

	CD Horizon	**CR Buffer**	**Freeze Horizon**	**Refresh Interval**	**CR Horizon**	**L1 Alert**	**L2 Alert**	**L3 Alert**
Baseline	180	180	40	20	360	180	130	30
Short	120	120	30	15	240	120	100	30
Long	240	240	70	40	480	240	160	30
Very Long	270	270	70	40	270	160	130	30

- Test Group 6 - Intruder Intent Change Stressor: Examined CD and CR performance when sudden traffic intent changes by the intruder result in CD well within AOP's detection horizon. The purpose of Test Group 6 was to provide data for an initial investigation of interoperability between strategic path planning systems such as FPM and tactical systems such as Airborne Collision Avoidance Systems (ACAS). Maneuvers were designed to represent a situation where a traffic aircraft changes its intent when in close proximity to the ownship, thereby creating a conflict. Although the FPM concept has a CP function to preclude such a situation, future operations may reveal it to occur occasionally. Because of a test design philosophy to substitute virtual test elements for live elements only when necessary, Test Group 6 required the intruder to downlink its change of intent to the ground station, which immediately uplinked the new intent to the ownship. As with Test Group 5, not enough flight test time was available to complete all planned maneuvers for Group 6. Two maneuvers were completed.

- Test Group 7 - Conflict Resolution and Prevention in Corridors Stressor: Exercised the ability of AOP's algorithms to allow aircraft to operate safely and maintain traffic flow within airspace corridors. AOP's current algorithms were designed for operations in open airspace. The open airspace may contain regions of restricted airspace that are small relative to the overall airspace. Flow corridors, with boundaries defined by large regions of restricted airspace, will likely be required for many UML-4 operating environments. Test Group 7 was designed to provide data for an initial exploration of UML-4 operations within corridors. The DFW UML-4 operating environment's corridors were utilized for Group 7. The corridors are necessary for UAM aircraft to access DFW terminals without encroaching on the surrounding runway operations, or to provide shortcuts through large regions of restricted airspace. Test Group 7 relied heavily on the virtual elements of the LVC test, as the regions of restricted airspace were much larger than regions of unrestricted airspace. Group 7 was made up of three maneuvers. All aircraft were required to comply with an RTA. For one maneuver, the ownship and the intruder were required to merge prior to entering the DFW terminal corridor. Their RTA point was located at the terminal.

- Test Group 8 - High Traffic Density Stressor: Explored FPM automation behaviors and functionality at traffic density levels expected for the higher end of UML-4. The purpose of Group 8 was to stress the AOP algorithm in several ways. High traffic levels directly impact the speed of the CR computation and the ability of the CR algorithm to compute resolution solutions. AOP's CP function, which prohibits solutions that create new conflicts, may also adversely impact the ability to determine CR solutions. Test Group 8 maneuvers modified a Group 1 maneuver by adding additional virtual traffic. Group 8's design goal was to double and triple virtual traffic density, but a limitation in the virtual traffic generator prevented density to increase more than 32 percent over the Group 1 baseline case. As currently implemented, the traffic generation capability requires every virtual aircraft to have an origin and destination. The simulated UML-4 environment's vertiports reached maximum capacity at the traffic limit used for Group 8. Two test maneuvers were executed.

To effectively meet the research purposes of each test run, a collection of distinct maneuvers was designed. A maneuver refers to a specific trajectory flown by the ownship and the intruder to create an intended encounter or event. For traffic conflict maneuvers involving both research aircraft, one of six encounter geometries were used. For a graphical representation of each of the six core encounter geometries, see Appendix B.3.

- Crossing: An opposing direction encounter where the trajectory of the ownship and the intruder crosses at approximately 120° while maintaining a safety margin of 0.2 NM.
- Acute Crossing: A same direction encounter where the trajectory of the ownship and the intruder crosses at approximately 30° while maintaining a safety margin of 0.2 NM.
- Head-On: An opposing direction encounter where the trajectory of the ownship and the intruder align toward one another with a safety margin of 0.2 NM.
- Intruder Overtake: A same direction encounter involving a significant speed difference where the intruder starts behind the ownship on a parallel trajectory, which includes a safety margin of 0.2 NM.
- Ownship Overtake: A same direction encounter involving a significant speed difference where the ownship starts behind the intruder on a parallel trajectory, which includes a safety margin of 0.2 NM.
- Merging: A same direction encounter where the trajectory of the ownship and the intruder merge together at a downstream location with a safety margin of 0.2 NM.

A set of 43 maneuvers was designed to conduct the research defined by each test group. A brief description of each maneuver is captured in Table 5. Maneuvers were constructed using a variation of an encounter geometry (if applicable) along with a combination of altitude changes, virtual traffic constraints, RTA change events, varying AOP time parameters, intruder intent change events, corridor constraints, and increased traffic density. Both the ownship and the intruder performance envelopes were represented by a modified version of the NASA quadrotor eVTOL model. The quadrotor model features a maximum airspeed of 109 kt, a minimum airspeed of 60 kt, and a nominal cruise CAS of 93 kt. For the purposes of the flight test, the minimum speed of the model was changed to 75 kt to work within the minimum CAS limits of both research aircraft. Execution procedures were determined by the associated test group objectives. Each maneuver underwent a verification process to ensure that it conformed to its specifications, including wind conformance, using batch simulations. Certain maneuvers were excluded from the test for various reasons defined in Appendix B. Removed maneuvers are presented with a strike through the row in Table 5.

7.2 Test Execution

Prior to flying each day of the two-week flight campaign, a safety briefing was held which covered expected environmental conditions and planned test operations. During a nominal flight test sortie, SARA and OPV flew 4D trajectories defining flight paths that would have resulted in the two aircraft losing separation. To begin the setup for a maneuver, the Middleware Engineer would designate an orbit location for both aircraft to head to and upload it to each aircraft. During this part of the setup, an onboard wind reading would be taken to be used in selecting which maneuver wind combination file most closely represented the current wind conditions at test altitude. Once in range, the pilots were commanded by the test conductor to initiate an action on their tablet that guided their aircraft into their designated orbit location. After the aircraft were established in their orbits, the pilots would be commanded by the test conductor to release from their orbits and fly toward their respective time-synchronized starting locations via automated guidance by the Middleware. After both aircraft crossed their respective T_0 point, the FPM maneuver began.

From there, AOP would continually monitor its 4D trajectory for conflicts with traffic, time of arrival constraints, or area hazards. Once AOP detected a conflict, it would notify the pilot of the conflict via the FPM Engineering UI and compute several resolution trajectories (up to four types: lateral, vertical, speed, and hybrid lateral-vertical-speed resolutions). The research pilot could select, review, and execute any of the provided resolutions depending on the current test group's procedures. If a resolution was executed, AOP then sent the executed 4D trajectory resolution to the AMM system, which commanded SARA to follow the updated 4D trajectory from the AOP software. A safety pilot ensured the safety of all executed trajectories.

The FPM Research Engineer was responsible for making the call to end each run, which was determined on a case-by-case basis depending on when all valuable data had been obtained. If the specific maneuver called for no execution of resolution advisories, the FPM Research Engineer would make the call to the test conductor to end the run at approximately 30 seconds until LOS. Furthermore, if a resolution advisory was executed, the FPM Research Engineer would make the call to end the run once the resolution was complete and the two aircraft had diverged by at least 0.5 NM. If a circumstance arose that an FPM Research Engineer decision could provide valuable data, and it was safe to do so, the FPM Research Engineer was allowed to make the end-run call when they saw fit. All decision making done by the FPM Research Engineer was aided using the FPM Ground Station display.

The flight test was conducted from October 16 to 26, 2023. During those two weeks, eight sorties were flown resulting in the completion of 34 FPM test points as shown in Table 6. All test systems and their functionalities performed sufficiently, resulting in the successful collection of FPM data. Additionally, there were no test days lost to adverse weather.

Table 5. FPM Maneuver Descriptions

ID	Test Group	AOP Time Parameters	Background Traffic	RTA Change	Intruder Intent Change	Description
M1	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Crossing, Level flight
M2	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Head-on, Level flight
M3	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Acute crossing, Level flight
M4	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Intruder overtake, Level flight
M5	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Ownship overtake, Level flight
M6	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Crossing, Intruder descends
M7	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Head-on, Intruder descends
M8	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Crossing, Intruder climbs
M9	1 & 2	Nominal	Nominal	None	None	Traffic conflict, Head-on, Intruder climbs
M10a	3	Nominal	Nominal	Delay	None	RTA compliance, Straight route, No intruder
M10b	3	Nominal	Nominal	Delay	None	RTA compliance, Straight route, No intruder, With execution
~~M11~~	~~3~~	~~Nominal~~	~~Nominal~~	~~Arrive Early~~	~~None~~	~~RTA compliance, Straight route, No intruder~~
M12	3	Nominal	Nominal	Delay	None	RTA compliance, Lateral route, No intruder
M13a	3	Nominal	Nominal	Arrive Early	None	RTA compliance, Lateral route, No intruder
M13b	3	Nominal	Nominal	Arrive Early	None	RTA compliance, Lateral route, No intruder
~~M14~~	~~3~~	~~Nominal~~	~~Nominal~~	~~Arrive Early~~	~~None~~	~~RTA compliance, Straight route, Virtual intruder~~
M15	4	Nominal	Nominal	Delay	None	Traffic conflict, RTA compliance, Intruder overtake
M16	4	Nominal	Nominal	Arrive Early	None	Traffic conflict, RTA compliance, Intruder overtake
M17	4	Nominal	Nominal	Delay	None	Traffic conflict, RTA compliance, Intruder overtake
~~M18~~	~~4~~	~~Nominal~~	~~Nominal~~	~~Arrive Early~~	~~None~~	~~Traffic conflict, RTA compliance, Intruder overtake~~
M19	5	Short	Nominal	None	None	Traffic conflict, Crossing
M20	5	Long	Nominal	None	None	Traffic conflict, Crossing
M21	5	Very Long	Nominal	None	None	Traffic conflict, Crossing
M22	5	Short	Nominal	None	None	Traffic conflict, Head-On
M23	5	Long	Nominal	None	None	Traffic conflict, Head-On
M24	5	Very Long	Nominal	None	None	Traffic conflict, Head-On
M25	5	Short	Nominal	None	None	Traffic conflict, Acute crossing
M26	5	Long	Nominal	None	None	Traffic conflict, Acute crossing
M27	5	Very Long	Nominal	None	None	Traffic conflict, Acute crossing
M28	5	Short	Nominal	None	None	Traffic conflict, Intruder overtake
M29	5	Long	Nominal	None	None	Traffic conflict, Intruder overtake
M30	5	Very Long	Nominal	None	None	Traffic conflict, Intruder overtake
M31	6	Nominal	Nominal	None	2 min until LOS	Traffic conflict, Crossing
M32	6	Nominal	Nominal	None	1 min until LOS	Traffic conflict, Crossing
M33	6	Nominal	Nominal	None	2 min until LOS	Traffic conflict, Head-on
M34	6	Nominal	Nominal	None	1 min until LOS	Traffic conflict, Head-on
M35	6	Nominal	Nominal	None	2 min until LOS	Traffic conflict, Acute crossing
M36	6	Nominal	Nominal	None	1 min until LOS	Traffic conflict, Acute crossing
M37	6	Nominal	Nominal	None	2 min until LOS	Traffic conflict, Intruder overtake
M38	6	Nominal	Nominal	None	1 min until LOS	Traffic conflict, Intruder overtake
M39	7	Nominal	Nominal	None	None	Traffic conflict, Merging, Corridor operations
M40	7	Nominal	Nominal	Delay	None	Traffic conflict, Intruder overtake, RTA Change, Corridor operations
M41	7	Nominal	Nominal	None	None	Traffic conflict, Head-on, Corridor Operations
M42	8	Nominal	1.32 Times	None	None	Traffic conflict, Crossing
M43	8	Long	1.32 Times	None	None	Traffic conflict, Crossing

Table 6. Executed FPM Runs During Flight Test

Maneuver ID	Test Group	Wind File	Resolution Execution	RTA Change	Time Parameters	Date
M1	1	00Kt-F000	No Execute	None	Nominal	10/18/2023
M2	2	00Kt-F000	Execute	None	Nominal	10/18/2023
M3	1	00Kt-F000	No Execute	None	Nominal	10/18/2023
M4	2	00Kt-F000	Execute	None	Nominal	10/18/2023
M5	1	00Kt-F000	No Execute	None	Nominal	10/18/2023
M6	2	00Kt-F000	Execute	None	Nominal	10/18/2023
M7	1	10Kt-F180	No Execute	None	Nominal	10/19/2023
M8	2	10Kt-F180	Execute	None	Nominal	10/19/2023
M9	1	10Kt-F180	No Execute	None	Nominal	10/19/2023
M15	4	10Kt-F180	Execute	Arrive Early	Nominal	10/19/2023
M16	4	10Kt-F180	Execute	Delay	Nominal	10/19/2023
M17	4	10Kt-F180	Execute	Arrive Early	Nominal	10/19/2023
M10a	3	10Kt-F300	No Execute	Delay	Nominal	10/23/2023
M13a	3	10Kt-F300	No Execute	Arrive Early	Nominal	10/23/2023
M13b	3	10Kt-F300	Execute	Arrive Early	Nominal	10/23/2023
M19	5	10Kt-F240	Execute	None	Short	10/24/2023
M20	5	10Kt-F240	Execute	None	Long	10/24/2023
M21	5	20Kt-F240	Execute	None	Very Long	10/24/2023
M31	6	10Kt-F240	Execute	None	Nominal	10/24/2023
M32	6	10Kt-F240	Execute	None	Nominal	10/24/2023
M39	7	10Kt-F180	Execute	None	Nominal	10/24/2023
M40	7	10Kt-F180	Execute	Delay	Nominal	10/24/2023
M41	7	10Kt-F180	Execute	None	Nominal	10/24/2023
M42	8	20Kt-F240	Execute	None	Nominal	10/25/2023
M43	8	20Kt-F240	Execute	None	Long	10/25/2023
M1	2	20Kt-F240	Execute	None	Nominal	10/25/2023
M2	1	20Kt-F240	No Execute	None	Nominal	10/25/2023
M3	2	20Kt-F240	Execute	None	Nominal	10/25/2023
M4	1	20Kt-F240	No Execute	None	Nominal	10/25/2023
M5	2	10Kt-F240	Execute	None	Nominal	10/25/2023
M6	1	20Kt-F300	No Execute	None	Nominal	10/26/2023
M8	1	20Kt-F300	No Execute	None	Nominal	10/26/2023
M22	5	20Kt-F300	Execute	None	Short	10/26/2023
M23	5	20Kt-F300	Execute	None	Long	10/26/2023

8 Results

8.1 Data Validity Analysis

After the flight test was completed, all data were examined to ensure they were valid for analysis. In-flight conformance with intended flight plans was a planned verification check, and is described in Section 8.1.1.

A software integration error that affected several test points was discovered during the post-test examination. An analysis was conducted to determine which test points were affected. Section 8.1.2 describes the error and the resulting validity analysis.

8.1.1 Trajectory Conformance

Since the FPM concept assumed some degree of conformance to a 4D flight plans, an analysis of 4D trajectory conformance was performed as one of the steps to establish the range of usable data from the flight test. Trajectory conformance buffer parameters within AOP were set based on the conformance capabilities of the test aircraft, and were identical for all background traffic aircraft. The SARA or OPV trajectory was considered to be in conformance when the measured deviations from the 4D flight plan were within the limits defined by the TPUBs. That is, the aircraft was in conformance if it passed each point of the 4D flight plan within 150 feet horizontal distance from the predicted latitude and longitude, ± 25 feet from the predicted altitude, and ± 3 seconds from the ETA at that point.

Trajectory conformance was measured by the FPM/AOP TPUBs Calibration Analysis Tool (FATCAT) software tool as described in Appendix E for both SARA and OPV. The procedures described in that appendix defined the time intervals in which usable data were collected. The first 10 seconds of each test run (starting at T_0) and the last 10 seconds before AOP was stopped were excluded.

With few exceptions, both aircraft stayed in conformance throughout each test run. Most of the exceptions indicated improvements that could be made in AOP's flight planning functions and the UI to AOP in order to enable the aircraft guidance to remain in conformance when executing new 4D flight plans from AOP.

Appendix E discusses specific cases of non-conformance with the 4D flight plan. Section 8.3.1 describes possible improvements to AOP that were judged likely to enable trajectory conformance to be maintained.

8.1.2 Scenario Conformance

In order for data to be useful, it was necessary for encounters between SARA and OPV (or between SARA and new RTAs) to occur as planned for each test run. All the desired encounters occurred, but the handling of the wind prediction during the flight test had some unintended effects on the observed results.

After the wind prediction for each run was determined as described in Section 6.4.1, the test design called for the wind prediction to be sent to AOP as part of its initialization sequence. A combination of multiple defects in multiple software modules, however, caused AOP to obtain its wind data from a source that specified a wind speed of zero for every run and caused this error not to be detected during system integration. The error was discovered during analysis of the results of the completed flight test.

This use of a zero-wind prediction in place of the selected wind prediction had the potential to affect AOP's behavior in many ways. These effects had to be accounted for when assessing the metrics of the flight test.

Because AOP produced both the initial active flight plan and the resolution flight plans during the flight test, both kinds of flight plan were affected equally. Therefore there were no cases of the "trivial" resolutions described in Section 6.4.2. The primary effects observed were due to the upper and lower constraints on the cruise speed imposed by the procedural limits of the aircraft model used for the ownship. Specifically, AOP was forbidden to plan a cruise airspeed less than 75 kt CAS or greater than 109 kt CAS.

In every case where the selected wind prediction had a non-zero wind speed, the predicted wind was a headwind or a crosswind and the actual airspeed required to meet an RTA or to maintain a planned ground speed was higher than AOP calculated. The following effects on airspeed might therefore have been expected to occur:

1. AOP might fail to slow the aircraft down to 75 kt CAS in test points that called for the minimum airspeed to be used at some point during the run.

2. AOP might direct the aircraft to exceed 109 kt CAS in test points that called for the maximum airspeed to be used at some point during the run.

3. AOP might direct the aircraft to exceed 109 kt CAS in test points that did not call for the maximum airspeed to be used at any point during the run.

4. The airspeed of the aircraft might change as it passed turns in the route, since AOP assumed a constant zero headwind but the actual headwind varied with the heading of the aircraft.

The last effect was judged to be irrelevant to the metrics of the flight test and was not examined. Every completed run was examined for the first three effects. Only the first two effects were found. The first effect occurred only when AOP created a flight plan with a calculated cruise airspeed of 75 kt CAS. The second effect occurred only when AOP created a flight plan with a calculated cruise airspeed of 109 kt CAS.

Table 7 describes the 10 test runs that were affected by AOP's use of zero wind. The maneuver number for each of these test runs appears on the top row of the table; the next row is filled if the research pilot was instructed to execute a resolution advisory, blank if the research pilot was instructed not to execute any resolution advisory. The next two rows show the predicted wind speed and direction from which the wind came in each test run. In each remaining row, a cell is filled if the effect described in the leftmost column occurred or may have occurred, empty if that effect was ruled out. The following is a description of each of these effects.

- **A missed RTA was predicted contrary to the test design.** The test point was designed to allow the aircraft to meet its initial RTA while flying at a speed greater than 75 kt CAS in the predicted wind. Due to using a zero wind speed, however, AOP predicted that it could not meet the RTA without decreasing the cruise airspeed below 75 kt CAS, and it therefore predicted that the initial flight plan would arrive too early at the RTA (a "missed RTA").

- **Flight plans likely missed an RTA by more than the test called for.** When attempting to adjust the cruise speed to avoid being early at an RTA, the minimum cruise speed selected by AOP resulted in a speed greater than 75 kt CAS in the predicted wind, causing AOP to predict an earlier arrival (a greater deviation from the RTA) than if it had used the predicted wind in its calculations.

- **Flight plans likely missed an RTA by less than the test called for.** When attempting to adjust the cruise speed to avoid being late at an RTA, the minimum cruise speed selected by AOP resulted in a speed greater than 109 kt CAS in the predicted wind, causing AOP to predict an earlier arrival (a smaller deviation from the RTA) than if it had used the predicted wind in its calculations.

- **RTA resolutions occurred before any RTA change or conflict.** Because AOP predicted a missed RTA in its initial active flight plan, it computed RTA resolutions starting shortly after passing the T_0 point, prior to any change in the RTA and prior to any conflict. No test points were designed to cause an RTA resolution under those circumstances.

- **RTA resolutions occurred after a CR.** In some cases AOP computed RTA resolutions after resolving a conflict, presumably because AOP could not find a maneuver that met the RTA while avoiding the conflict. It is possible for this to occur even when AOP uses predicted wind as intended, but it could also happen because the missing wind prediction causes AOP to calculate too high a ground speed on the flight path that avoids the conflict.

- **The predicted conflict location may have been shifted.** By design, the flight plans predicted by AOP for the modeled eVTOL used a lower airspeed in descent than in cruise. The use of zero wind speed rather than the predicted wind speed in AOP's calculations would mean that the actual cruise airspeed was reduced from what the test design intended while the descent airspeed was increased, thereby causing the aircraft to arrive at the planned conflict location late. This discrepancy in timing could have caused AOP to predict the planned conflict at a time and location different from the time and location planned in the test design. (In an extreme case, it is conceivable that by arriving late, AOP

Table 7. Test Points with Results Affected or Possibly Affected by Wind Prediction Errors

Maneuver number	M1	M2	M3	M4	M10a	M13a	M13b	M21	M42	M43
Pilot execute resolution	yes	no	yes	no	yes	yes	yes	yes	yes	yes
Wind speed (kt)	20	20	20	20	10	10	10	20	20	20
Wind from direction (deg)	240	240	240	240	300	300	300	240	240	240
A missed RTA was predicted contrary to the test design	•	•	•	•				•	•	•
Flight plans likely missed an RTA by more than the test called for					•					
Flight plans likely missed an RTA by less than the test called for						•	•			
RTA resolutions occurred before any RTA change or conflict	•	•	•	•				•	•	•
RTA resolutions occurred after a CR								•		•
The predicted conflict location may have been shifted	•	•	•	•				•	•	•
Resolution flight plans might have been "more extreme" or "less extreme" than if AOP had used the predicted wind					•					
Flight plans may have exceeded 109 kt CAS actual airspeed						•	•			

could have missed the planned conflict altogether, but no such event was observed during the flight test.)

- **Resolution flight plans might have been "more extreme" or "less extreme" than if AOP had used the predicted wind.** A resolution route that would have been accepted by AOP, taking the predicted headwind into account, might have been rejected during the flight test because AOP, using zero wind speed, calculated that the airspeed would be too low. The resolution route finally accepted in the flight test might use a larger path deviation (a "more extreme" maneuver) in order to meet the RTA without using excessive speed according to AOP's calculation.

  Conversely, an RTA resolution that would have been rejected by AOP due to excessive airspeed, taking the predicted headwind into account, might have been accepted during the flight test due to AOP using zero wind speed in its calculations. A successful resolution of the same RTA, taking the predicted headwind into account, would likely have required a greater reduction in the length of the flight path; hence the length reduction that actually occurred would have been smaller (a "less extreme" maneuver).

- **Flight plans may have exceeded 109 kt CAS actual airspeed.** A flight plan generated by AOP using zero wind speed in its calculations that assumed a cruise airspeed of 109 kt CAS (the upper limit) would have caused SARA to exceed 109 kt CAS if SARA followed that flight plan in a headwind. (Alternatively, a flight plan that assumed a cruise airspeed near 109 kt CAS also could have caused SARA to exceed 109 kt CAS due to headwind, but analysis of the actual data ruled out this possibility in all of the completed test runs.)

Appendix F provides a more detailed description of the interaction of wind effects with every test run.

Aside from these effects, the scenarios in the completed runs progressed essentially as planned. The principal effects observable in the test results are RTA resolutions observed but not executed for some test runs where no RTA resolutions were expected, and some cases in which an unmet RTA may be due solely to the mishandling of the predicted wind.

8.2 Function Verification Analysis

The following sections present results for Element 1, Function Verification. MOPs results are presented in Section 8.2.1. Success criteria results and quantitative results are provided based on the metrics. Additional context is provided as needed. Section 8.2.2 provides an overall analysis of MOPs findings, and Section 8.2.3 relates the MOPs findings to the MOEs.

8.2.1 Functional Performance Results

- MOP 1: There is a low fraction of false conflict alerts (false positives) and a very low fraction of missed conflict alerts (false negatives).
  - Criteria

    i. Missed detection (false negative) rate of less than 5%.

    ii. False detection (false positive) rate of less than 20%.
  - Result: Successful

    i. There were no false negative detections.

    ii. The false positive detection rate was 0.92%.

- Context

  The total number of CD test points for all test runs was 4529. Performance was measured for all CD cycles. The highly successful results are likely due to the test design assumption that all aircraft conform with high precision to ground-referenced 4D trajectories. All aircraft in the test complied with their shared intent. Note that results do not reflect generalized performance because the test design may introduce biases. Each maneuver was designed to create a conflict with the intruder.

- MOP 2: Conflicts are detected within the first two CD cycles, measured from when the first detection should have occurred.

  - Criterion

    Detection occurs within the first two CD cycles for 90% or more occurrences.

  - Result: Successful

    100% of conflicts are detected within the first two cycles.

  - Context

    The total number of conflicts detected across all test points was 34. The initial CD time was measured against the starting time of CD lookahead horizon or, when applicable, from the first time of a triggering event that occurred within the CD horizon.

- MOP 3: AOP provides at least one CR advisory for each detected conflict within the first CR refresh cycle.

  - Criterion

    90% or more of all conflicts result in a resolution being offered within the first CR cycle.

  - Result: Successful

    97% of conflicts result in at least one resolution advisory across all CR cycles. For the 97% of resolvable conflicts, a resolution is provided within the first CR cycle for 100% of occurrences.

  - Context

    Only 1 out of 34 conflicts never results in a resolution advisory. The unresolvable conflict occurs during test maneuver M40, involving corridor operations. The test run exhibited a flickering location of predicted LOS. Refer to Section 8.4.11 for a discussion of the failed resolution occurrence.

- MOP 4: Over the design range of CR validity, executed resolutions result in no LOS between the resolving aircraft and any conflicting aircraft.

  - Criteria

    i. 90% or more of all executed resolutions satisfy minimum separation standards at CPA.

    ii. 100% of all executed resolutions are flyable within the hosting aircraft's performance limits.

  - Result: Successful

    i. 100% of executed resolutions do not result in a LOS with the corresponding conflicting aircraft.

    ii. 100% of executed resolution trajectories are entirely within the ownship's performance constraints and other constraints imposed by test procedures.

- Context

  The total number of executed resolutions in the analysis set was 15. For conflicts that were resolved with vertical separation, the average separation at CPA was 501 ft, with a standard deviation of 2 ft. For conflicts that were resolved with lateral separation, the average separation was 2960 ft, with a standard deviation of 316 ft. The maximum lateral separation at CPA in the data set was 3375 ft, and the minimum was 2543 ft. The minimum lateral separation expected from a AOP computation is 1800 ft, made up of the 1500 ft separation minimum and a 0.05 NM (304 ft) resolution buffer added by AOP. Given the average aircraft speeds in the test, combined lateral and along-path TPUBs can result in more than 1000 feet of additional separation, depending on the encounter geometry.

  All separations at CPA fall within expectations. AOP's CR algorithm seeks to provide an optimal resolution solution, which drives separation values toward minimum separation criteria. The algorithm must also account for all other traffic, which may result in higher separations with the original conflicting aircraft.

  Two maneuvers in the analysis set involved climbing or descending encounters. Maneuver M6 was resolved with a hybrid solution that provided lateral separation. That maneuver was added to the lateral separations analysis data set. Maneuver M8, resolved with a vertical solution, provided more than 830 ft of vertical separation. It was therefore treated as an outlier and deleted from the vertical separations analysis data set.

- MOP 5: Conflict resolutions do not result in a repeat conflict between the resolving aircraft and the original conflict aircraft within the design range of CR validity.

  - Criterion

    90% or more of all executed resolutions do not result in a second conflict with the original conflicting aircraft.

  - Result: Successful

    100% of executed resolutions do not result in repeat conflict with the original intruder aircraft.

  - Context

    A repeat conflict is defined as a conflict that occurs with a traffic aircraft following the resolution of a previous conflict with the same aircraft. Of the 21 executed resolutions, there were no occurrences of repeat conflicts with the design range of CR validity. There was a single occurrence of a repeat conflict during a test maneuver that involved corridor operations. The second conflict's time to first LOS (TTFLOS) was predicted to occur 379 seconds after the first conflict's resolution was executed, beyond the CR lookahead horizon used for the run.

- MOP 6: Conflict resolutions do not create new traffic or airspace conflicts within the design range of CR validity.

  - Criterion

    90% or more of all executed resolutions do not result in additional conflicts with traffic or area hazards in the surrounding environment.

  - Result: Successful

    100% of executed resolutions do not result in a LOS with the intruder or virtual background traffic.

- Context

  Of the 21 executed resolutions, the were no occurrences of a new conflict with any traffic within the range of CR validity. For the baseline CD and CR horizon settings, it is possible for a new conflict to be detected when the ownship reaches the point of first LOS with the original conflict. The CR horizon is exactly twice the CD horizon. While such an occurrence is acceptable based on the technology's design, it did not occur in the flight test.

- MOP 7: Conflict resolutions meet an RTA constraint within aircraft performance limits.

  - Criterion

    90% or more of all CRs will arrive within three seconds before or after the RTA, assuming the CR is executed at least three minutes prior to arriving at the destination.

  - Result: Successful

    92.2% of traffic CRs meet the RTA within the RTA tolerance.

  - Context

    All maneuvers were evaluated as a combined set, except for Maneuvers M10a and M17. M10a was omitted because it had a conflict with a virtual traffic aircraft near the destination due to the use of incorrect wind (see Section 8.1.2). By design, AOP's attempt to resolve this conflict (which was unresolvable due to the proximity to the destination) created dozens of failed one-second CR cycles. This behavior was considered invalid for this analysis. Maneuver M17 was omitted because it was designed to force relaxation of RTA compliance. A total of 103 CR trajectories made up the analysis set. All resolution trajectories in the analysis set were computed before three minutes prior to arrival at the RTA.

    There was at least one AOP resolution advisory in each CR refresh cycle that met the RTA within three seconds for 92.2% of occurrences. For all AOP resolution advisories (each refresh cycle containing up to four, corresponding to the four degrees of freedom), the RTA was met within tolerances for 87.5% of occurrences. Note that AOP is designed to always prioritize resolving traffic conflicts over complying with its assigned RTA.

    All test maneuvers were designed for the ownship's trajectory to comply with its assigned RTA prior to encountering a traffic conflict or an RTA change. Because of an error that prevented AOP from reading current winds information, this design goal was not met for some maneuvers. For these maneuvers, AOP detected RTA noncompliance based on its assumption of zero winds. If possible, AOP then provided a resolution advisory that would meet the RTA without violating its speed constraints. If not possible, AOP provided the best non-compliant solution, which involved a trajectory flying at a speed limit. If AOP was able to comply with the RTA, under some circumstances the updated trajectory required the ownship to fly near the aircraft's high or low speed constraint. A subsequent conflict can cause a resolution trajectory to be limited by a speed constraint, thereby preventing RTA compliance. Although the measured result meets the success criterion, the percentage is therefore probably lower than it would have been if the AOP winds data acquisition error was not present. Note that due to the incorrect wind information AOP used, the CAS limits AOP imposed did not correspond to the actual airspeed limits.

- MOP 8: FPM technology prioritizes CRs over RTA constraints. If necessary to resolve a traffic conflict, it provides a CR solution that is out of compliance with the RTA, but as close to compliance as possible.

- Criteria

  i. Conflicts are successfully resolved for 90% or more of all occurrences.

  ii. RTA constraints are optimally relaxed to meet the hosting aircraft performance constraints for 90% or more of all occurrences.

- Result: Successful

  i. 100% of conflicts presenting potential prioritization of RTA are resolved.

  ii. 100% of traffic CRs minimally relax RTA constraints.

- Context

  M17 was the only test maneuver intentionally designed to force relaxation of RTA compliance to resolve the conflict. AOP performed as designed, relaxing RTA compliance to successfully resolve the conflict with the intruder. The ownship arrived 17 seconds early by speeding up to resolve a conflict with the overtaking intruder, resulting in an arrival time that was 14 seconds out of tolerance. The CR's predicted CPA with the intruder was 3590 ft laterally. An analysis of AOP's data records for several PBGA iteration generations provides evidence that the RTA was relaxed to a minimum extent. All resolution candidates in the speed-only population that arrived 17 seconds early or earlier were conflict-free. Those that arrived later conflicted with the intruder. Interestingly, even though minimum RTA relaxation was achieved, results suggest the resolution solution could have been further optimized. As the PBGA fitness function is currently designed, no other optimization criteria (such as predicted battery usage or the magnitudes of the speed changes for the maneuver) were considered when the RTA was not met. See Section 8.4.10.

• MOP 9: FPM technology resolves traffic conflicts in airspace corridors of width between 0.5 and 1 NM that are not constrained by RTAs located within the corridor, at expected UML-4 density.

- Criteria

  i. Conflicts involving corridor operations are resolvable for 90% or more occurrences.

  ii. Executed resolutions meet or exceed separation standard minimums at CPA for 90% or more of all detected hazards.

- Result: Unsuccessful

  i. 75% of all conflicts relating to corridor operations are resolvable.

  ii. 100% of executed resolutions meet minimum separation standards at CPA.

- Context

  Three test maneuvers involved corridor operations. One maneuver had two conflicts, for a total of four conflicts. The conflict of maneuver M40 was unresolvable due to a continuously flickering location of the point of separation loss between the hosting aircraft and the intruder. In maneuver M39, a second conflict occurred beyond the range of CR validity for the first conflict. The point of separation loss for the second conflict was predicted to occur during the landing phase of the intruder, which was in front of the ownship. Both aircraft were scheduled to arrive at the same vertiport. The RTAs of the two vehicles were 29 seconds apart, guaranteeing a conflict.

• MOP 10: FPM technology completes all CR computations and displays CR alerts within the first 15 seconds of each refresh cycle at UML-4 traffic density.

  - Criterion

    The computation and presentation of resolution advisories to the operating pilot is achieved within 15 seconds for 100% of all CR cycles.

  - Result: Successful

    100% of all CR cycles are computed and presented to the pilot within 15 seconds.

  - Context

    The analysis set contained 213 CR cycles. See Section 8.3.3 for additional discussion.

- MOP 11: FPM technology provides operators with sufficient time to evaluate and execute a resolution before a new set of resolution advisories become available to evaluate.

  - Criterion

    The operating pilot is provided at least 15 seconds to review the latest computed set of CRs for at least 90% of all CR cycles that produced at least one valid resolution.

  - Result: Successful

    95.2% of all resolution sets were provided to the pilot for at least 15 seconds before being removed or replaced.

  - Context

    A total of 124 CR cycles were analyzed. All cycles in the analysis set contained baseline time parameters and provided at least one valid resolution for display to the pilot. The analysis set was also restricted to cycles displayed long enough for the resolution end to be bounded by the start of another CR cycle. Display windows are bounded by the start and stop times of two CR cycles. If a pilot executes a resolution, there is not a subsequent CR cycle that can be used to determine the end of the display preview time. The baseline CR refresh cycle of 20 seconds was sufficient to provide at least 15 seconds of resolution display time for 95.2% of all occurrences.

8.2.2 Functional Performance Analysis

Based on test results, ten of the eleven measures of performance met their success criteria. The CD function performed well, with no missed conflict alerts and a low number of false conflict alerts. The reference technology always detected conflicts within six seconds of the earliest time the conflict could have been detected. The technology was able to provide high-quality resolution advisories for 97% of all occurrences, and these were provided in the first resolution cycle. Resolution trajectories were always flyable within the limits of aircraft performance. No new conflicts were created by the resolution trajectories within the CR lookahead horizon, and there were no repeat conflicts with the original conflicting aircraft. CRs met an RTA within three seconds for 92% of all occurrences, and traffic CRs were successfully prioritized over RTA compliance. RTA compliance was relaxed to a minimum extent necessary. All CR computations were completed within 15 seconds, even for high-density traffic scenarios. Pilots had at least 15 seconds to evaluate resolution advisories for 95% of all occurrences. The single unsuccessful MOP involved operations within airspace flow corridors. Only three of the four conflicts that occurred within corridors were resolved by the reference technology. As expected, test data sets were too small to allow for statistical analysis. The results can only provide an initial indication of success. The MOPs and success criteria were designed to have future application in traffic simulation experiments, which can generate larger and more comprehensive data sets.

8.2.3 Effectiveness Analysis

Flight test results indicate MOE 1 (vehicle-centric flight path management) is feasible. Using the reference technology, pilots were able to manage their individual flight paths in UML-4 traffic density. They avoided other traffic (MOE 1a) and regions of restricted airspace (MOE 1b), and had they flown the complete planned trajectory, they would have arrived on time (MOE 1c). They always operated within the performance constraints and procedural constraints of their aircraft. These conclusions are supported by the results of MOPs 1 through 8.

Judging by the limited data of the flight test, MOE 2 (operational scalability) also appears to be feasible. The reference technology had adequate computational speed to handle UML-4 traffic density, as indicated by MOP 10. The technology does not currently filter traffic, so it was detecting conflicts and providing resolutions for all traffic in the test. Higher traffic density scenarios presented no problem for the technology. MOP 11 results showed that if a CR refresh cycle of 20 seconds or greater is found to be adequate, pilots will have at least 15 seconds to evaluate the automation's advisories before selecting and executing. The evaluation pilots also provided feedback regarding a future human/machine interface (see Section 8.3.6). Their comments suggest that with the inclusion of a mature interface, a 15-second evaluation time will be acceptable. These results suggest MOE 2a is met. Test results also provided no indication of a domino effect (MOE 2b). Resolutions did not cause repeat conflicts, as shown by MOP 5 results. The CP function performed well, with resolution trajectories producing no new conflicts (MOP 6 results). The domino effect and overall system stability will depend on many factors, such as CR lookahead horizon settings and the possibility that traffic aircraft will need to update their intent often. These issues must be evaluated in future traffic simulation studies. While conclusions about overall system stability are limited based on the flight test, results indicate the reference automation technology can reliably produce conflict-free trajectories while resolving conflicts.

MOE 3 (technology maturity) appears to be feasible for the open airspace of the test. The open airspace contained area restrictions, with unusable airspace making up 11.6% of the total. Based on the results of MOPs 1 through 8, 10, and 11, the computing power was fully adequate for operations in this airspace. Results from MOP 9 suggest it is premature to draw conclusions about MOE 3 for corridor operations. Sophistication of the technology will probably need to increase, so future computing requirements are unknown. Nevertheless, the short compute times achieved using today's low-cost computing platforms strongly suggest computing power will not be a limiting issue. It should be noted that computing platform certification factors and any additional computing burdens associated with collaborative and responsible automation are not considered in this conclusion.

MOP 9 provided results for MOE 4 (operations in highly restricted airspace). MOP 9 did not meet one of its success criteria. Flight test results therefore do not indicate operations within corridors are feasible using the reference technology. To reduce the challenge, testing was restricted to scenarios only having an RTA that was located outside the airspace corridors. Even those were only partially successful. Corridors may require a form of automation based on relative time, such as in-trail spacing, rather than the absolute-time form used by the reference technology. Prior to future research in support of MOE 4, the reference technology should be refined to address the specific challenges of highly restricted airspace.

8.3 Concept and Technology Advancement Analysis

The following sections present results for Element 2, Concept and Technology Advancement. These analyses explore early-stage considerations that may inform the development of future functional requirements, with rationale grounded in observed trends from the gathered flight test data. While not intended as formal guidance, the findings contribute to a growing understanding of the challenges and operational needs anticipated under airspace conditions with higher-density and greater complexity.

8.3.1 Trajectory Conformance

The analysis of trajectory conformance in Section 8.1.1 and Appendix E revealed several ways in which the support for 4D conformance provided by AOP might be improved. This included cases in which the aircraft did not lose 4D conformance in the sense that it did not go outside the bounds of the TPUBs. The following improvements to AOP were judged most likely to improve trajectory conformance.

8.3.1.1 Better Enforcement of the Freeze Horizon in PBGA

The "freeze horizon" is a configurable parameter that PBGA uses when creating candidate 4D trajectories to resolve a conflict or to meet an RTA. The 4D trajectories created by PBGA start at the time when the PBGA request was initiated, but the aircraft typically progresses along its existing flight plan while PBGA is computing its resolution and while the air crew is evaluating and executing a resolution.

A maneuver by PBGA that started immediately would likely produce a trajectory that had diverged from the existing trajectory by the time the new 4D flight plan was executed, causing the aircraft to be instantly out of conformance with its new 4D flight plan. The purpose of the freeze horizon is to ensure that a sufficiently long initial portion of the new 4D trajectory coincides well enough with the existing 4D trajectory to allow the new 4D flight plan to be executed without a loss of conformance.

The algorithms that enforce the freeze horizon worked well for some maneuvers, but careful examination of the flight test results showed some cases in which the initial part of the new 4D flight plan did not match the existing 4D flight plan as closely as desired.

8.3.1.2 Applying a Freeze Horizon for Speed when the RTA Changes

In the AOP function that computes a new 4D flight plan to meet a new RTA (if possible) by changing the cruise airspeed, the trajectory is computed as if the aircraft set a new cruise airspeed at the instant the new RTA was received. In the flight test, since the aircraft continued to fly at the old cruise airspeed until the air crew executed the new 4D flight plan, the along-path error of the aircraft's position tended to increase suddenly when that flight plan was executed. While several test points ran successfully without a loss of 4D conformance despite experiencing this phenomenon, loss of conformance is possible.

8.3.1.3 Applying a Freeze Horizon when Lateral or Vertical Changes Affect Speed

A similar effect sometimes occurred due to a lateral resolution or the lateral component of a hybrid lateral-vertical resolution, when a change in the length of the remaining horizontal path required a significant change in the cruise airspeed in order to continue to meet the RTA. In this case, as in the previous case of a change in cruise airspeed, AOP based its new flight plans on the assumption of a speed change at the instant PBGA started to compute the resolution, leading to a sudden increase in along-path error when the new flight plan was executed.

8.3.1.4 Expiration of Advisories in the User Interface

At some point, each new 4D flight plan proposed by PBGA must diverge from the 4D flight plan that the aircraft is already following. It is desirable to have a mechanism to detect when the aircraft has flown past the point at which the aircraft can correctly implement a particular 4D flight plan and remove the ability of the air crew to execute that 4D flight plan.

This was a feature of some previous flight replanning systems supported by AOP, but it has only been applied to the maneuver patterns that were developed for the Trajectory Aware Planner (TAP) (Ref. 24) and was not available in the flight test.

8.3.2 Conflict Detection

An analysis provided estimates of the number of true positive conflicts, false positive conflicts, true negative conflicts, and false negative conflicts that occurred over the course of the flight test. For some estimates, an analysis data set was made up of specific runs that were designed to provide the needed information. For other estimates, the outcomes of each CD cycle were combined into an analysis set. An AOP detection cycle occurred every three seconds. Details are provided in Appendix A.

Results are shown in the form of an error or confusion matrix in Table 8. A total of 4529 CD cycles resulted from 41 test points. There were 34 conflicts detected during the test, and 1082 CD cycles were associated with those conflicts. TPUBs are used to drive the number of false negatives (missed alerts) low, potentially at the expense of increased false positives (false alerts). For the TPUBs settings used in the test, less than one percent of false positives were observed. There were no false negatives. The overall prediction accuracy was determined to be 99.78%.

Table 8. Conflict Detection Error Matrix

CD Cycles	Predicted True	Predicted False		
Actual True	True Positives 1072	False Negatives 0 (Type II error)	Precision 100%	False Omission 0%
Actual False	False Positives 10 (Type 1 Error)	True Negatives 3447	False Discovery 0.29%	Negative Prediction 99.71%
	Sensitivity 99.08% (True Pos. Rate)	Misses 0% (False Neg. Rate)	Prediction Accuracy 99.78%	
	Fall-Out 0.92% (False Pos. Rate)	Specificity 100% (True Neg. Rate)		

The test design should be kept in mind when viewing these results. All aircraft in the test were flying a ground-referenced 4D trajectory with high precision, and all aircraft conformed to their shared intent. TPUBs values were set based on the observed conformance precision. Conflicts occurred by design for almost all test points. Encounters between the ownship and the intruder were designed to have 0.2 NM lateral offsets for safety.

8.3.3 Conflict Resolution

Flight test results were analyzed to characterize performance of the CR function. Analyses of Section 8.2 found that resolutions, when provided, are flyable by the ownship and are effective in achieving separation. In routine operations, time must be available for detection and resolution of conflicts from the point of first detection to the point of transition to tactical separation or collision avoidance. If human decision makers are in the loop, plenty of time must be available for them to evaluate and select resolutions that best meet mission objectives, even in the presence of system delays associated with ground-based decision-making. A wide time range of reliable CR may therefore be necessary. Untested factors, such as how far in advance an aircraft can predict its intent and how well it can maintain that shared plan, may influence effectiveness of CR well before a LOS occurs. However, the current test may provide insights regarding the rest of the time range, up to and including resolutions shortly before LOS. Performance characteristics of interest include:

- availability of resolution solutions as the aircraft in conflict approach one another
- effectiveness of each of the four resolution degrees of freedom
- performance for encounters that involve an aircraft climbing or descending, and
- performance consistency across a range of lateral encounter angles.

The analysis data set was generated from all test runs that involved CD and CR. Runs that involved pilot execution of a CR advisory generally produced only one set of resolution advisories from the first CR refresh cycle. Most data were therefore contributed by test runs that did not involve pilot execution. For these runs, up to four resolution advisories were generated from each CR refresh cycle, and refresh cycles were repeated until the run was ended. The complete data set was made up of 111 traffic CR cycles from all 34 maneuver test points. The six test points with time parameter changes from the baseline are included in the total set. Any resolutions associated with incomplete CR cycles are not included in the analysis set. Incomplete cycles are cycles that were terminated by AOP before running for the full duration, and can result from several sources: pilots executing a resolution provided from the previous cycle, the conflict no longer being detected, or a CR restart triggered by a new conflict or RTA noncompliance situation being detected.

8.3.3.1 Time to First Loss Analysis

The number of resolutions available for display in each refresh cycle is expected to decrease with the time available before LOS. Resolution maneuvers can require small changes from the original plan when distances between conflicting aircraft are large. Adjustment magnitudes must increase as the aircraft get closer, so aircraft performance limits may become a factor. Maneuvering area may also decrease.

Results are normalized as average percentages because sample sizes varied across the time range. If all four degrees of freedom were available for all runs in a cycle, the y-axis value would correspond to 100% for that cycle. The width of each bar corresponds to the duration of the CR resolution cycle. The time of resolution is associated with the start of a cycle, so the first resolutions are shown just prior to the CD horizon of 180 seconds. (Conflicts were observed to occur a few seconds prior to the detection horizon, and the CR cycle is triggered almost immediately after detection.) All CRs in the data set occurred no later than about 40 seconds prior to first LOS. This was probably caused by termination of the test runs. CR would normally continue to 30 seconds, the point of transition to tactical hazard avoidance. Baseline alert levels associated with priority rules are shown with extended tick marks on the x-axis. A Level 1 alert (180 seconds) corresponds to a point-out for awareness if the ownship has right-of-way. A Level 2 alert (130 seconds) corresponds to the time the ownship operator should act if the intruder has not resolved the conflict. Alert levels were not used during the flight test because only the ownship was equipped with AOP, but alert levels are important when considering availability of resolutions.

An average of about three resolutions are available (75% availability) from 180 to 90 seconds TTFLOS, although availability drops to under 70% around 140 seconds, followed by a recovery to over 80%. The variations may be caused by the small sample size, or there may be an unidentified cause. As expected, a decline is noticeable as TTFLOS decreases, beginning around 80 seconds. The decline occurs within the

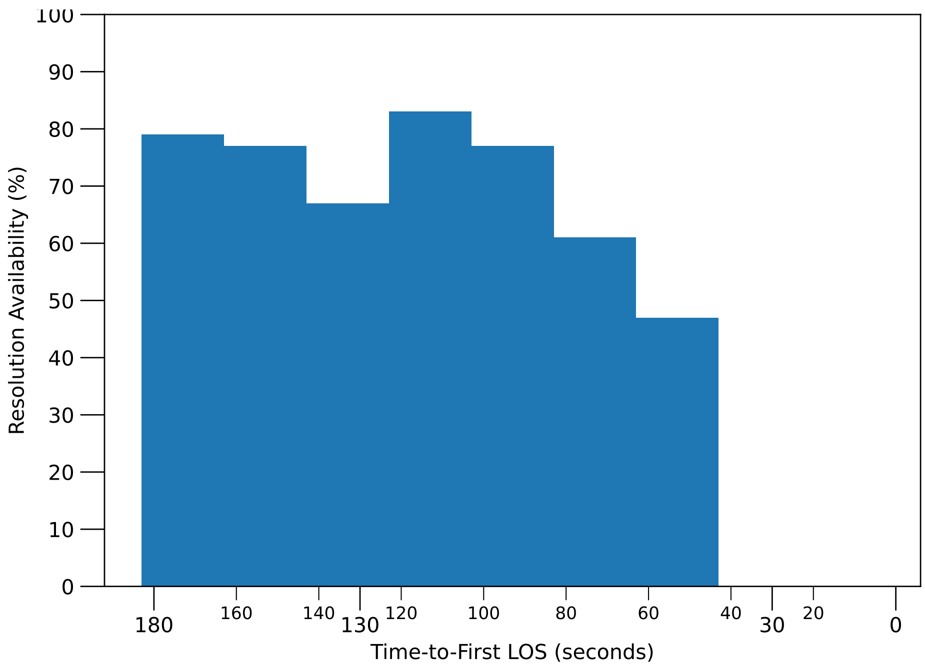


Figure 11. Resolution Availability as an average fraction of resolutions available as a function of time to first LOS (TTFLOS) for all runs that used baseline time parameter.

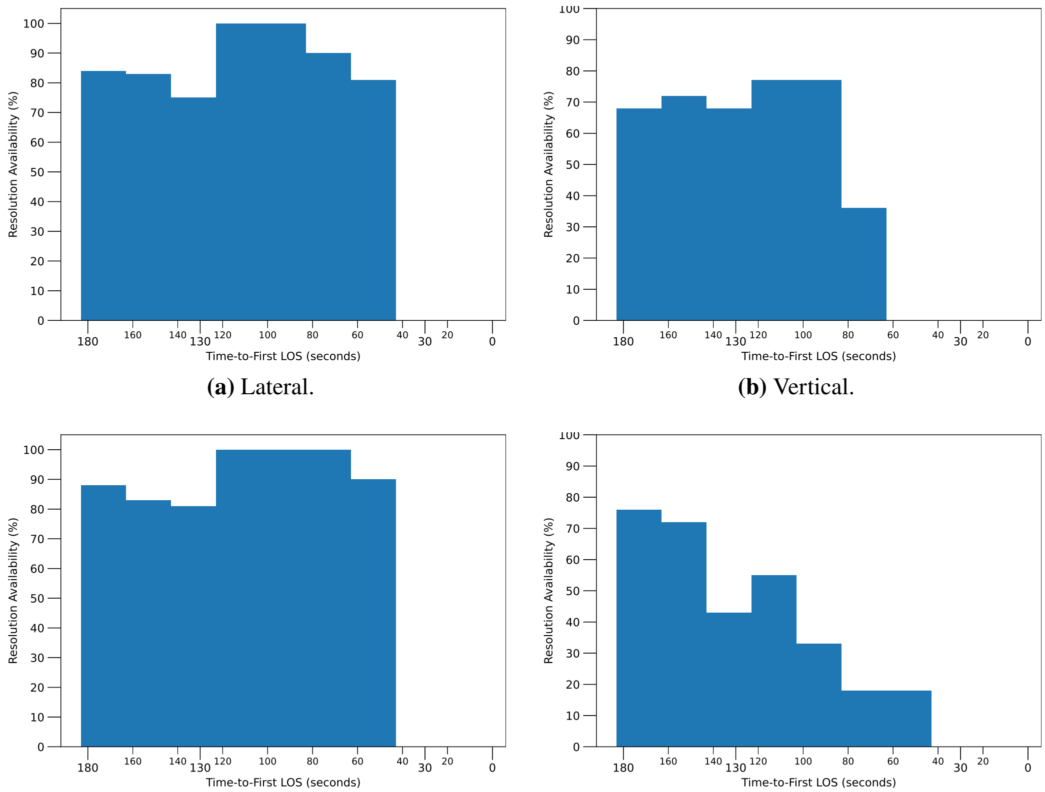
Level 2 alert period. Further refinement of the technology may allow for future availability increases in this time range. If that is not possible, it may be beneficial to increase time horizons to allow for Level 2 alerting to begin earlier than 130 seconds TTFLOS. It should be noted that only one resolution is necessary. Multiple resolution advisories are offered to provide optimal solutions and accommodate user preferences.

The availability of each of the four resolution degrees of freedom were determined from test results, and are shown in Figure 12. Results are again normalized as percentages. For each cycle, the fraction of runs containing a solution corresponding to the plot's degree of freedom is shown on the y-axis. The lateral and hybrid populations maintain high fractions of resolution advisories across the entire range, with the range of 120 to 80 seconds showing 100% resolution availability for both degrees of freedom. The vertical populations had lower availability, with a sharp fall-off at roughly 80 seconds TTFLOS. The vertical degree of freedom may be impacted by climb or descent rates at close ranges. The speed populations show a near-steady decline in availability as TTFLOS decreases. The speed range of the ownship was constrained by test procedures, and PBGA limited solutions accordingly. The maximum and minimum allowed CASs of 109 kt and 75 kt significantly reduced the effectiveness of the speed degree of freedom at close ranges.

8.3.3.2 Conflict Encounter Analysis

Cycle-to-cycle resolution availability trends are shown in Figure 13 for the tested encounter incidence angles and vertical incidence geometries. Figure 13a shows the level flight runs for different incidence angles, Figure 13b shows results for intruder-in-climb conflicts, and Figure 13c shows results for intruder-in-descent conflicts. For all figures, the number of resolutions available is shown as a function of TTFLOS. Each run is shown by a different color line. Triangle symbols show the point of first resolution. Symbols without a trend line were either executed or aborted by the pilot before a second set of advisories was displayed.

Head-on, acute crossing, ownship overtake runs appeared to perform best, with four resolutions available for a significant portion of the time range. One of the three intruder overtake runs also produced four resolution advisories for most of the range. Crossing cases were mixed, with most producing three advisories. The two



(a) Lateral.

(b) Vertical.

(c) Hybrid lateral-vertical.

(d) Speed-only.

Figure 12. Resolution Availability as a Function of Time to Loss by Resolution Type

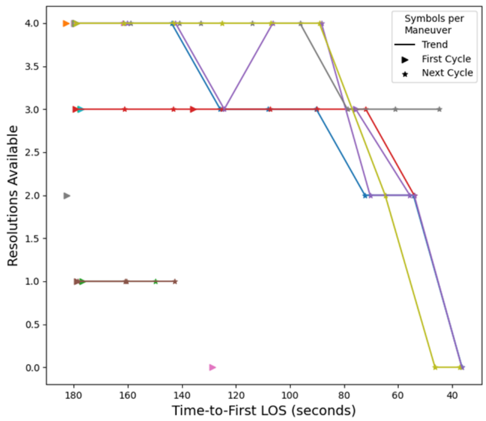
other intruder overtake cases performed worst overall, with the overtake run within a corridor producing no resolutions. Figure 13a shows most level-flight runs maintained the number of resolutions available until a reduction around 90 seconds TTFLOS. One level-flight run experienced a reduction of available resolutions in a cycle followed by an increase in the next cycle. Similar behavior is shown for one intruder climb and one intruder descent run in Figures 13b and 13c. All of these behaviors are in the range of 130 seconds TTFLOS, suggesting there is an unidentified cause. All results may be impacted by specifics of the maneuver designs and therefore may not indicate behavior trends.

8.3.3.3 Resolution Failure Analysis

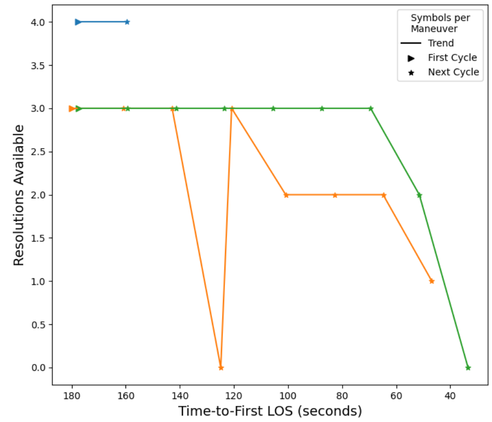
A CR cycle failure (a non-resolving PBGA cycle) is defined as any completed cycle that does not provide at least one resolution advisory. Out of the 111 total traffic CR cycles, AOP was unable to provide a resolution for eight cycles. The overall resolution cycle failure rate for traffic conflicts is 7.21%.

Of the eight cycles without resolution, six (75%) occurred within 50 seconds of TTFLOS. Within that subset, five out of six represent the last CR cycle before 30 seconds TTFLOS, the point of tactical override. Only one instance of two sequential CR cycles where no resolutions were found occurred (Run M3, Group 1). In one instance (Run M8, Group 1), a CR cycle provided no resolution advisory where both the previous and following CR cycles did.

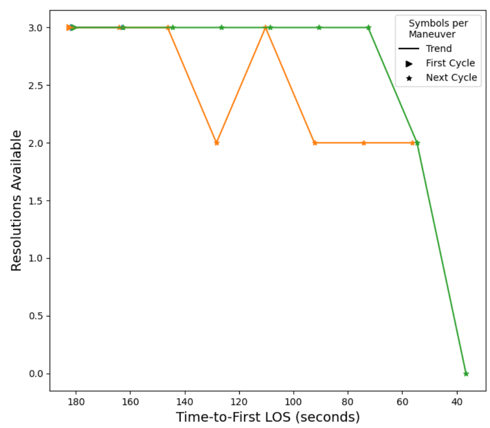
Resolution failure rates are shown in Table 9 for the full analysis set. The speed degree of freedom is least successful in providing resolutions, followed by the vertical degree of freedom. In addition to the TTFLOS impacts mentioned above, resolution failure rates may be affected by the number of available patterns for



(a) Level.



(b) Climb.



(c) Descent.

Figure 13. Resolution Trends by Encounter Type

these populations. Lateral and hybrid degrees of freedom each have three patterns, whereas only one pattern exists for the speed and vertical types.

Table 9. Resolution failure rates by degree of freedom across the TTFLOS range.

Degree of Freedom	Failure rate (%)
Lateral	18.02
Vertical	40.54
Speed	49.55
Hybrid	14.41

Resolution degrees of freedom selected for execution by the research pilots are shown in Table 10. Pilots selected the vertical and speed advisories least often, despite one of the pilots suggesting he prefers vertical resolutions because they are easy to understand. Both pilots had difficulty with the FPM Engineering UI profile display, which displayed vertical and speed resolution advisories. Nevertheless, the lower availability of these types may have contributed to a reduction in their selection by the pilots.

Table 10. Pilot execution selections by degree of freedom.

Degree of Freedom	Executed Resolutions
Lateral	8 (38.1%)
Vertical	3 (14.3%)
Speed	3 (14.3%)
Hybrid	7 (33.3%)

8.3.4 Conflict Resolution Compute Time

Flight test results were used to evaluate PBGA computational performance on the AOP-hosting NUC processor (NUC11TNHi7, 64GB DDR4 Mem, 1TB PCIe M.2 SSD). The following sections provide a cursory look at times elapsed between a requested PBGA computation and its completion for display to the pilot. The analysis data set contained a total of 213 CR cycles across 33 runs. M10a was omitted from the data set for the reasons it was omitted from the MOP 7 analysis in Section 8.2.1.

8.3.4.1 Computation Time

A distribution of elapsed computing times for all CR cycles is shown in Figure 14. The average processing time for each PBGA resolution advisory is 5.41 seconds. The longest resolution solution required more than 11 seconds, and the shortest about 1 second. There is a peak in counts near the average, but a greater peak is observed in the range of 3 seconds, potentially indicating a bimodal distribution. Each degree of freedom was evaluated separately to gain further insight. Vertical and speed solutions were found to require less time than lateral and hybrid lateral-vertical solutions. This may be due to the time required to compute fitness penalties for conflicts, which are more complex when a lateral degree of freedom is involved. Average times were 2.74 seconds for vertical, 3.87 seconds for speed, 4.55 seconds for lateral, and 5.01 seconds for hybrid solutions.

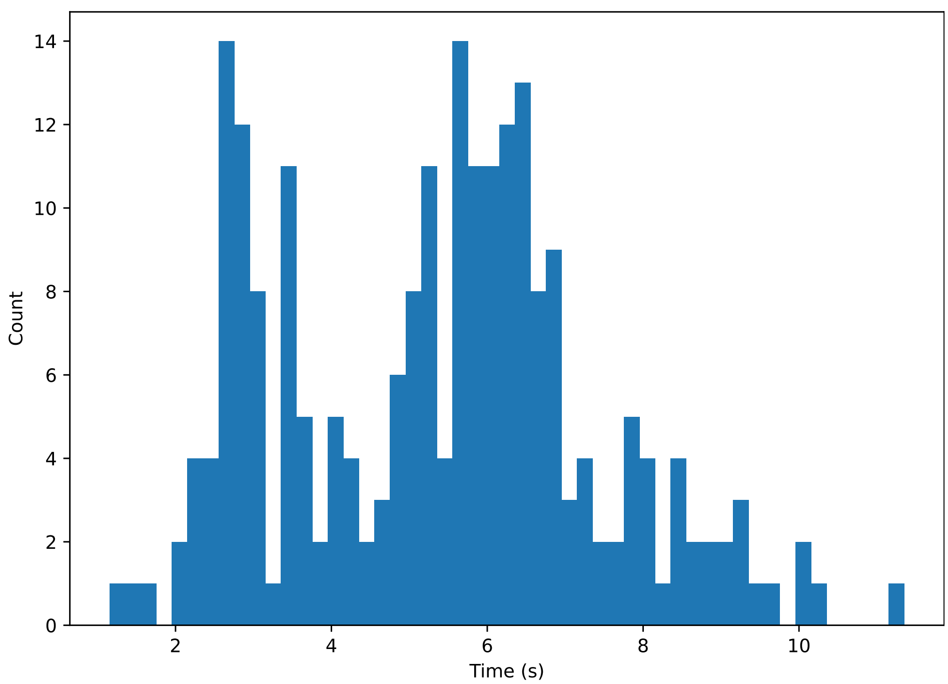


Figure 14. Resolution compute times for all CR cycles.

8.3.4.2 Time to First Loss Analysis

Figure 15 maps PBGA CR compute times as a function of the TTFLOS. Results in the figure represent PBGA cycles from all maneuvers, including maneuvers with non-baseline time parameters. Points shown beyond 200 seconds on the x-axis represent resolution cycles from maneuvers having long and very long time parameter settings. The figure shows little or no correlation.

8.3.4.3 Traffic Number Analysis

AOP was configured to perform no filtering of traffic during the flight test, so every aircraft in a test scenario was considered by the PBGA algorithm as it performed its function. Previous simulation results (Ref. 25) have shown an increase in traffic density causes an increase in PBGA computing time. For each CR cycle, the number of traffic aircraft in the scenario was recorded at the time of the cycle's start. Other live aircraft operating in the airspace were intentionally filtered to ensure a "sterile" research environment and therefore are not included for analysis here.

Figure 16 shows how the PBGA compute time was related to the number of traffic in the scenario each time PBGA ran to completion. Traffic aircraft (on the x-axis) include the real intruder and the virtual background traffic.

These data did not reproduce the relationship between number of traffic aircraft and PBGA compute time that was clearly observed in Reference 25. There are several possible causes for this. The most obvious reason is that almost all runs had numbers of traffic aircraft within a very narrow range of values, so the effect on the elapsed time would be relatively small, while there is considerable variance in the elapsed time at any given number of traffic (as Reference 25 also found).

Several other conditions likely contributed to the difficulty of finding a relationship between number of traffic and PBGA compute time in the flight test data. In the data collected for Reference 25, there were 6944 distinct ownship runs to consider (compared to 34 in the flight test); only the first execution of PBGA in each run was considered (because multiple compute times in the same run might not be independent of each other); the initial conditions and flight plans of the ownship and all traffic aircraft were randomized rather than planning for specific encounters; PBGA ran on a much older generation of hardware than the flight test used; the populations for the various degrees of freedom were computed sequentially on one thread, not concurrently on four threads; and there were never more than 70 traffic aircraft known to AOP. Notably, the mean PBGA compute time for 70 traffic aircraft in Reference 25 was longer than the longest PBGA compute time observed in the flight test for any number of aircraft.

A future study of the factors that influence PBGA compute time could look at a large number of randomized trials performed in simulation. It could consider the number of traffic aircraft; the number of area hazards as well as their complexity (measured by the number of sides of the hazard polygons); the number of waypoints in the active route at the start of PBGA; and whether the active route meets its RTA.

Further improvements likely could be made to PBGA compute time without changing hardware. In particular, there is no attempt at load-balancing within PBGA; each population is allocated one thread and is expected to run to completion on that thread. Typically some threads will be underutilized and the overall PBGA compute time will be determined by whichever population takes the most time to compute. The trend in computers is toward larger numbers of central processing unit (CPU) cores with multiple threads per core, so even at the beginning, PBGA is likely underutilizing available resources. But the essential structure of PBGA, in which each generation of a population requires the independent evaluations of multiple candidate flight plans, appears to be highly parallelizable. It seems likely that merely by making better use of threads, PBGA could handle much more complex airspaces in elapsed times similar to those observed in the flight test.

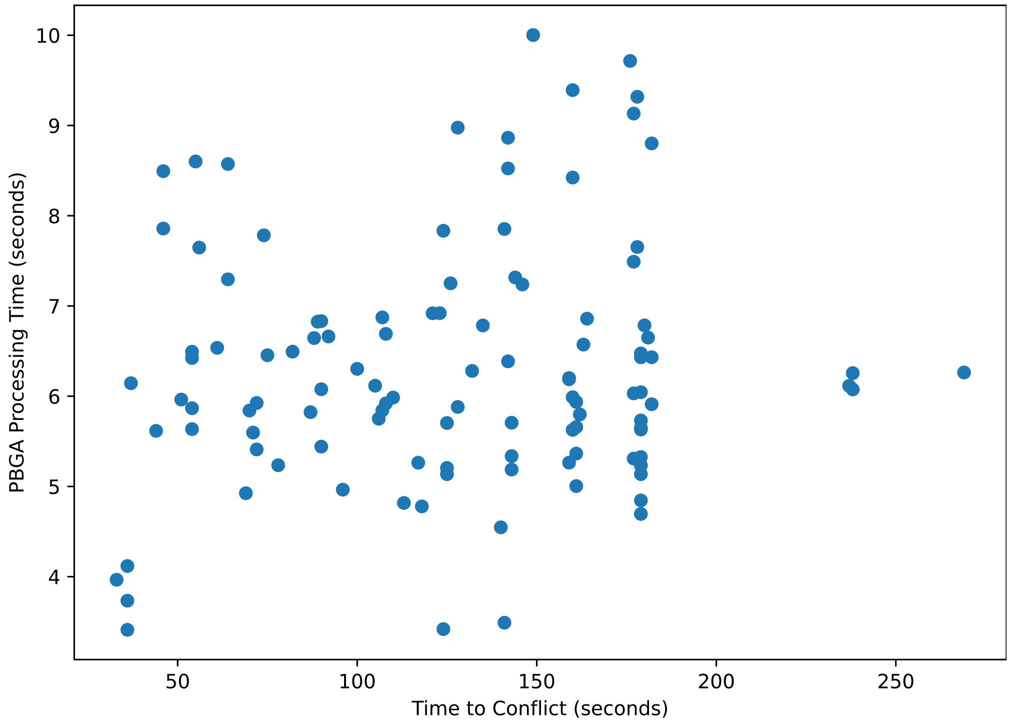


Figure 15. Impacts of TTFLOS on PBGA performance.

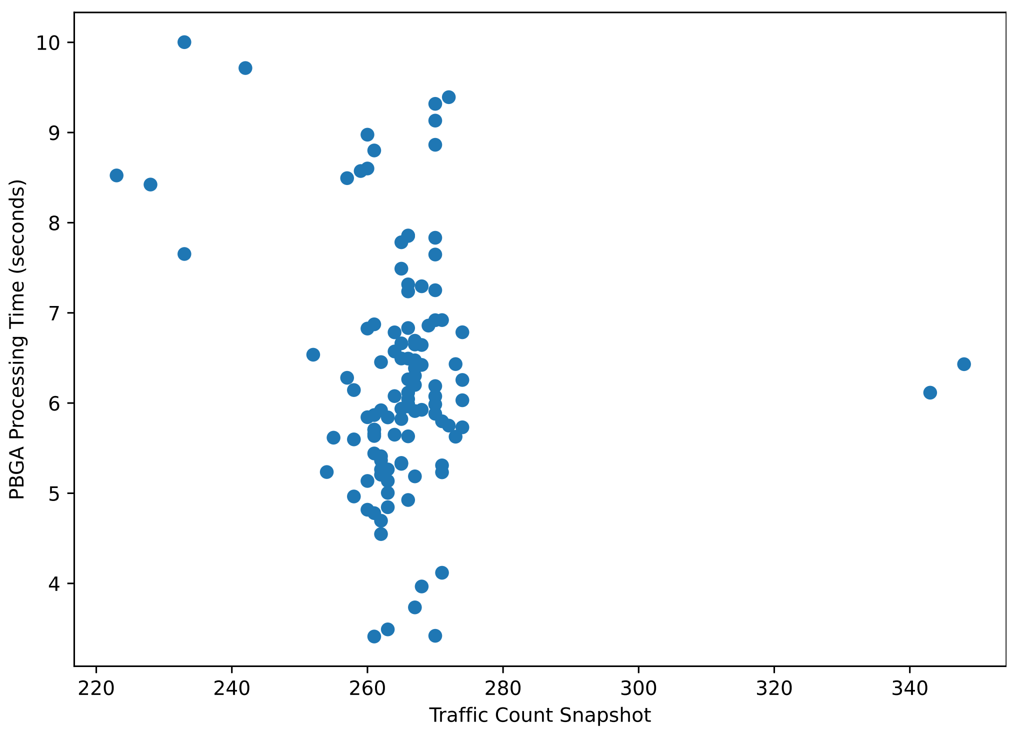


Figure 16. Impacts of traffic number on PBGA performance; includes CR cycles from all 34 runs.

8.3.4.4 Resolution Advisory Type Analysis

Three distinct groupings were analyzed to observe for any appreciable performance differences between the full set of PBGA cycles. Figure **??** shows PBGA performance for the three groups. Group 1 represents only cycles that consist of a hybrid resolution, Group 2 represents only cycles where all four advisory types provided a resolution, and Group 3 represents only cycles that consist of either a speed or vertical resolution. Groups 1 and 2 are found to be insignificantly different from that of the full data set. However, Group 3 is observed to skew toward the upper interquartile range, which consists of a significantly higher count of cycles with longer computation times.

8.3.5 Arrival Time Compliance

The Element 2 Arrival Time Compliance analysis provided an evaluation of AOP's ability to meet RTAs. The analysis focused on investigating RTA conformance as a function of time-until-RTA and RTA conformance separated by advisory population. Only resolutions that occurred at least three minutes prior to the RTA point were considered for this analysis (time-to-destination cut-off), but no resolutions were affected by this cut-off criterion. AOP was configured to consider a 4D flight plan in conformance with the RTA if the ETA at the RTA point was within ± 3 seconds of the specified constraint time (that is, up to three seconds earlier or three seconds later). Finally, all flight test runs were represented in the analysis except for M10a and M17 runs. M10a was excluded for the reasons it was omitted from the MOP 7 analysis in Section 8.2.1. M17 was excluded because that particular maneuver was designed to intentionally cause a traffic conflict that required AOP to relax its RTA conformance to avoid the aircraft.

Figure 18 shows RTA conformance for all PBGA advisories as a function of time. The majority of advisories conformed to AOP's acceptable RTA tolerance. Figure 18 also shows some advisories creating late arrivals, which was expected, and many advisories creating early arrivals, which was unexpected.

There were several cases where the advisory was predicted to arrive late at the RTA point. M13a (10kt-300° wind file) accounts for the majority of these positive ("late") RTA errors; six of the nine "late" advisories shown in Figure 18 occurred in M13a, all in the last three refresh cycles. M13a is an RTA delay maneuver, and the procedure required the research pilots to not execute a AOP advisory for the duration of the run. It was observed that each of the last three refresh cycles for M13a fell progressively further out of RTA conformance as the time-to-destination cut-off was approached. The interpretation of this behavior is that AOP was no longer able to generate a resolution advisory capable of meeting the new RTA because the ownship was too close to its destination and performance limits to acceptably absorb the delay. AOP instead produced an advisory that would deliver the ownship to the destination as close to the RTA time as possible. As stated earlier, this claim is further supported by the progressive conformance error observed between the refresh cycles as time-to-destination progresses. M13a was also a maneuver that was flown with incorrect winds, but the magnitude of wind error was low, so it is not expected to have significantly affected the run.

A software defect in AOP that was discovered during the analysis created several other advisories that did not conform to their RTAs. Logic that determined whether to call PBGA for RTA resolution and logic within PBGA that determined whether a resolution candidate was in conformance with the RTA both used a six-second tolerance rather than the configured three-second tolerance. As a result, in some cases PBGA delivered sets of advisories that were all out of conformance. For example, in M43, due to the use of a zero wind prediction in AOP instead of the prediction that was sent (20 kt from 240 degrees), the active flight plan at T_0 was predicted to arrive 7.8 seconds early at the RTA point; AOP therefore performed several RTA resolutions and one SICR in which PBGA converged to lateral and hybrid resolutions that arrived approximately 5.9 seconds early. (These resolutions were within the six-second tolerance; the reasons for

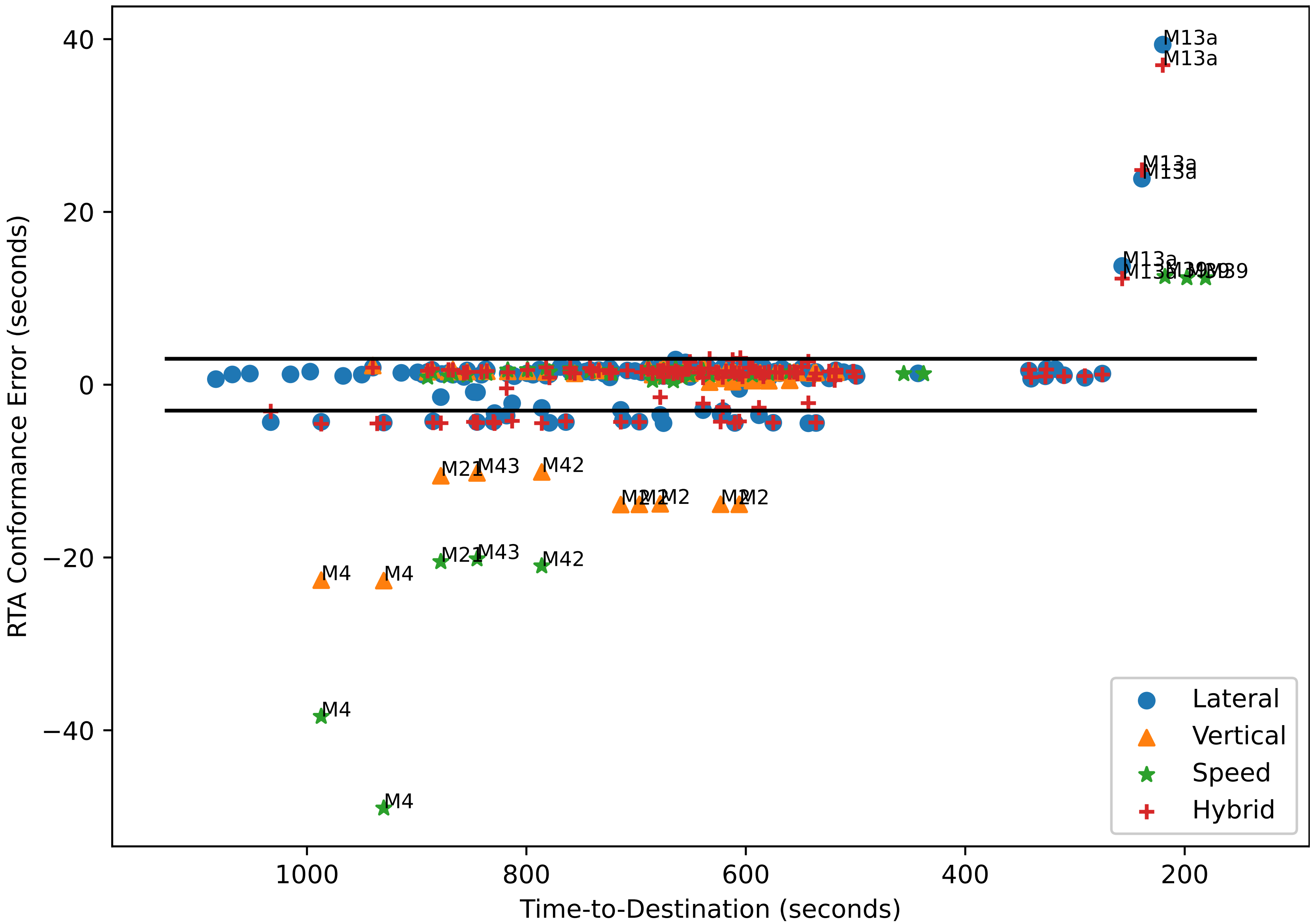


Figure 18. RTA Conformance Error of All Advisories vs. Time-to-Destination

the bias toward the low end of the tolerance interval have not been fully analyzed.) Another function in AOP was responsible for filtering advisories so that if any advisories were in RTA conformance, no out-of-conformance advisories would be displayed. The filtering function was correctly configured for a three-second RTA tolerance, so in cases where some advisories were 5.9 seconds early and the others were even earlier, it determined that all advisories were out of tolerance and passed all advisories to the engineering UI.

It appears likely that if AOP had been consistently configured to use the three-second tolerance in all logic that concerned RTA conformance, the lateral and hybrid advisories with times just below the lower conformance limit in Figure 18 would have been replaced by advisories that were in true RTA conformance at the cost of slightly greater energy usage, and the vertical and speed advisories further below the line would have been filtered out altogether.

M4 (20kt-240° wind file) showed large RTA conformance error for speed and vertical degrees of freedom compared to other runs for two particular refresh cycles between 1000 and 900 seconds time-to-destination. M21 (10kt-180° wind file), M42 (20kt-240° wind file), M43 (20kt-240° wind file), and M2 (20kt-240° wind file) also experienced significant RTA conformance error for speed and vertical resolution degrees of freedom, with speed producing the worst RTA conformance error for most cases. A major component of the conformance error observed in these specific cases was determined to be due to the predicted wind error experienced during most of the flight test. M42, M43, M2, and M4 experienced 20 knots of wind error, which would have had a considerable effect on AOP's ability to conform to the RTA. The failure to fully configure PBGA to the three-second RTA also likely contributed. Furthermore, earlier executions of M4 and M2, where the wind file matched the predicted wind error, performed within acceptable RTA conformance.

Table 11 provides a breakdown of RTA conformance for all advisories. The percentages shown are strongly impacted by the maneuver design as well as the software defect previously mentioned and should not be interpreted as generalizable results. The lateral degree of freedom conformed to its RTA at a higher rate, followed by vertical, speed and hybrid.

Table 11. RTA Conformance for All Advisories

Advisories	Total Advisories	Within RTA Tolerance	Rate of RTA Conformance
Lateral	141	117	82.98%
Vertical	58	48	82.76%
Speed	42	34	80.95%
Hybrid	115	90	78.26%
Total	356	289	81.18%

8.3.6 Pilot Assessment

8.3.6.1 Objective

Evaluation pilots were asked to provide subjective assessments of the reference automation and the feasibility of UML-4 operations for the distributed-agent FPM implementation under study. Pilot feedback was used to gain insights and draw initial conclusions related to the foundational research questions. Human/machine interface assessments and other assessments related to human factors were not a primary research focus. A broader evaluation of human factors related to the flight test is available in Monk et al (Ref. 26). Here we provide subjective feedback from the pilots at the conclusion of the flight test activity, specific to their experience with the AOP technology.

8.3.6.2 Method

Two NASA research pilots conducted all flight test runs in the ownship, supported by Sikorsky safety pilots. The research pilots were provided with subjective survey questions. Some questions were presented to the pilots on a tablet at the end of each maneuver. Feedback for the rest was received during a post-test interview. Post-test interviews were conducted with each pilot independently. The pilots were provided with questions prior to the test and were allowed to ask clarification questions during pre-test training. Pilots were provided with hard copies of the questions so they could make notes throughout the testing.

The following questions were presented on the tablet at the end of each maneuver. Feedback was received based on a rating scale, as follows: 1: strongly disagree; 2: disagree; 3: neither agree nor disagree; 4: agree; 5: strongly agree.

1. I had enough time to evaluate all resolutions while they were available.
2. All presented trajectory resolutions seemed safe.
3. All presented trajectory resolutions seemed feasible within the aircraft's performance envelope.

Research pilots provided responses to the following four questions at the conclusion of all testing. Feedback was received based on the above rating scale. Research pilots provided their responses in a debriefing at the end of the final test day.

1. I understand AOP's functions and behaviors.
2. I was never confused by AOP's alerts and advisories during flight.
3. I never felt a need for additional information to understand what AOP was advising.
4. All presented trajectory resolutions seemed reasonable for use in routine operations.

   The following survey questions requested feedback in the form of narrative responses. Research pilots provided their responses in a debriefing at the end of the final test day.

5. Please provide additional comments if you answered "disagree" or "strongly disagree" to any of the rating questions.
6. Were you less comfortable operating with AOP for certain test runs than others? If so, which test runs and why?
7. Which resolution type(s) did you tend to feel least comfortable with? Why?
8. What operational situations or flight factors would influence your prioritization of available options in resolving conflicts and meeting RTAs?
9. What additional information would you like AOP to provide, if any, to help you prioritize and select from the available resolution options?
10. Based on your experiences with AOP, do you have any suggestions for improving the trajectory solutions AOP provides?

    The following survey questions were added after all training was complete, so pilots did not have written copies of the questions. Research pilots provided their responses in a debriefing at the end of the final test day.

11. What are your thoughts about the separation standards we assumed for UML-4 and used in this test? (1500 ft laterally, 450 ft vertically)

12. The traffic you experienced is modeled to be representative of a future UML-4 operational environment. Do you have any thoughts about operating routinely in such an environment?

13. Do you have thoughts about whether UML-4 operations could safely be conducted in all meteorological conditions?

14. Do you have any recommendations for what is needed in future testing?

15. Additional time horizon questions:

    (a) Under any of the time parameters variation runs, did you feel you had more time than necessary? For instance, did AOP alert you to a conflict earlier than you felt necessary?

    (b) Do you have any specific comments about our Group six maneuvers, where the intruder changed its intent and created a conflict at two minutes and one minute to first LOS?

16. The only situational awareness you got for virtual elements (background traffic, SUA) was through the FPM engineering UI. Do you believe that SA was compelling enough to impact your decision-making during a run?

8.3.6.3 Training

All pilots received familiarization training. The NASA research pilots were provided with additional detailed training covering the FPM concept, the automation, operating procedures, and the specific objectives of each test group. Pilot training consisted of a dedicated simulation experiment conducted in the Fall of 2022, computer-based training with a desktop interface to a cloud-based simulation, and a dedicated crew training exercise held three weeks prior to the flight test. The HITL portion of the FPM-2 experiment, described in Section 5.2, served as an initial introduction to the automation and the FPM Engineering UI.

In addition to familiarization, the experiment allowed the pilots to provide feedback for design of the UI. This feedback was used to refine the FPM Engineering UI prior to the flight test. The computer-based training application enabled pilots to experience test maneuvers and operate AOP to detect and resolve conflicts. The application used a remote interface connected to a pilot-in-loop simulation consisting of an AOP-equipped aircraft operating in the full UML-4 traffic environment. The application was available to pilots for several months prior to the flight test. The final crew training exercise allowed the pilots to gain a full understanding of each test group and increase their proficiency with AOP. Pilots were made familiar with the pilot feedback questions to be asked at the end of each run and at the end of testing. The training consisted of a 3.5-hour briefing, followed by practice simulation runs in the NASA Armstrong Flight Research Center System Integration Laboratory. The System Integration Laboratory used hardware to host AOP and Middleware that was identical to test aircraft hardware.

8.3.6.4 Results

The following are high-level results obtained from the pilot feedback. Condensed transcriptions of the pilot interviews are presented in Appendix D.

- **Operator understanding of the FPM automation's functions and behaviors**

Both pilots felt that they understood the functions of the automation system. The FPM Engineering UI did not impede their understanding of the functions the automation was performing, but it did impede their ability to make best use of the automation. As pilots gained familiarity of the UI through use, their confusion was somewhat reduced, but both pilots requested additional information be displayed that would provide increased situational awareness. They highlighted specific issues and provided suggestions for the design of an operational human/machine interface. Details follow.

For the flight test, the FPM Engineering UI was designed to present a resolution advisory for each of the four degrees of freedom, with no ranking or recommendation to the pilot regarding which advisory to select. For each degree of freedom, the resolution solution with the best fitness score was presented regardless of its score relative to solutions for the other degrees of freedom. This created challenges for both pilots by increasing the time needed to evaluate and select an advisory. One pilot stated that the default time for selection of approximately 20 seconds was not enough to evaluate each advisory. He stated 40 seconds (one of the Test Group 5 parameter values) was probably enough time. The other pilot made similar comments, but added that if only one resolution was provided, he would probably only need five seconds for evaluation. Both pilots said they would prefer the automation provide a single best solution recommendation. Other possibilities they proposed were to have the automation withhold presentation of advisories that provided little difference from other slightly better advisories, and to have the automation rank advisories and present them one at a time, starting with the top-scoring advisory.

Both pilots stated the FPM Engineering UI's presentation of information hindered their ability to understand and evaluate advisories. They stated that the lateral degree of freedom was the easiest to interpret, and the speed-only degree of freedom was the most difficult. Both pilots noted several challenges in using the profile display, which presented information for vertical and speed advisories. One pilot requested a design that provides better situational awareness about time and distance to the point of conflict, especially for the profile display. He suggested turn points on the profile display would help. One pilot expressed a need to have awareness of the time available to select and execute an advisory. He also said he needs better awareness of the time before LOS occurs than was available in the FPM Engineering UI. He suggested the display continue to provide its information regarding the conflicting aircraft even after the conflict is resolved by the automation. Such information would help him monitor the situation to verify that the conflict has been resolved.

- **Quality of advisories provided by the FPM automation**

  Both pilots provided a rating of 4 ("agree") to the statement, "All presented trajectory resolutions seemed reasonable for use in routine operations." All maneuvers that made up the trajectory solutions were considered in the normal range, and were within the limits of the aircraft's performance. One pilot commented that he never saw a resolution advisory that he would "never do." Trust in the provided resolutions by operators was emphasized by both pilots as being critical for an operational system.

  Neither pilot was fully comfortable with the advisories provided by AOP. One pilot mentioned that the most intuitive advisories for him were vertical resolutions. The other pilot mentioned lateral resolutions. Both pilots expressed difficulty with speed-only resolutions, but both mentioned their lack of comfort may have been a result of the current FPM Engineering UI display design.

  Both pilots acknowledged that the automation system had more information than they did about traffic and hazards in the environment, thereby creating resolutions that were not always intuitive to the operator. Based on their comments, one pilot may have been willing to rely on trust that a mature system would utilize the information properly, while the other pilot requested more information be provided, allowing him greater situational awareness. Specific counterintuitive advisory characteristics noted by the pilots were:

- Hybrid resolutions and vertical resolutions were often similar. Most hybrid solutions were made up of a similar vertical component with the addition of a small lateral component. The pilots' observations were likely caused by the fact that an independent advisory for each degree of freedom was presented in the test. No advisory was suppressed because it provided no benefit over the others.
- One pilot observed a resolution where the ownship climbed while the intruder was already at a higher altitude. The ownship appeared to be climbing into the intruder. While the intruder's intent may have indicated it would be descending, the pilot considered the advisory counterintuitive.
- Lateral resolutions sometimes began a turn-back maneuver earlier than the operator expected, such as before the ownship was abeam of the intruder when the two aircraft were approaching head-on.
- One pilot observed that resolution advisories sometimes differed significantly from one refresh cycle set to the next, causing him to wonder why, and resulting in him requesting an interface that would provide him more situational awareness.
- One pilot mentioned some resolution advisories crossed the flight path of the conflicting aircraft, which he would prefer not to do. When asked whether an ellipsoid shape of the protected zone could be preferable to the current hockey puck shape, both pilots stated that having additional separation when passing in front of another aircraft would increase their comfort.

- **Operations within airspace corridors**

  Operating within flow corridors was acknowledged by both pilots as being more challenging than operating in open airspace having a few airspace constraints. One pilot felt "more constrained" in corridors, and he openly wondered whether he would be able to receive acceptable quality resolution advisories within corridors. He suggested CR advisories would need more sophisticated algorithms to be acceptable for use in corridors. The other pilot noted that corridors will need to be designed to allow resolutions to occur within them. He also noted hazardous weather could disrupt operations within corridors.

- **Anticipated utility of FPM automation in a future UML-4 operating environment**

  In addition to CD and CR, both pilots said there would be value in conserving onboard energy for eVTOL aircraft, so trajectory optimization is of high value. Neither pilot mentioned a capability to precisely comply with an assigned RTA as an important function. Both pilots also indirectly suggested additional information could be used by the algorithm to provide resolution advisories tailored to pilot preferences. As an example, one pilot mentioned that if he was flying at an altitude with good visibility, he would prefer to not select an advisory that would cause him to change altitude and fly into clouds. The other pilot provided the example of a need to avoid turbulence. As a general comment, he said his preferred resolutions depend on the situation he is in. For example, he would prefer not to climb if he is near his destination. He also mentioned he may have personal preferences, such as preferring turns over climbs, which suggests there may be value in adding a capability for the automation to accept individual preference inputs. When asked if passenger comfort would influence his preferences, one pilot mentioned he might prefer resolutions that provide greater separation from traffic and resolutions that are less aggressive, possibly causing him to reject speed-only advisories.

- **FPM feasibility in a UML-4 operating environment**

  The pilot assessments suggest that FPM automation or a similar function would be necessary in a future UML-4 operating environment. With a collaborative and responsible FPM function, routine operations at UML-4 seem feasible to the pilots. If an operator is in the loop, pilot feedback indicates the automation's advisories must be fully trusted to detect and resolve conflicts, with no expectation that the operator will verify correctness of an advisory. One pilot commented that in a visual environment, the 1500 ft separation standard assumed for the test seemed large enough, but he cautioned that he could not comment about operating in instrument meteorological conditions. He stated he would need to experience that to have an opinion. He would prefer to have larger buffers surrounding the protected zones, but he also said pilots may be able to get comfortable with small separation standards. The other pilot made similar comments. He expressed concern for overall system stability as traffic density increases. He noted shared intent might change too frequently at high traffic density for separation assurance to be effective. He also mentioned that at high traffic densities, display of relevant traffic information to the pilot may be difficult, resulting in loss of situational awareness. He recommended the UI be designed to present only the information needed by the pilot. He asked for only the information he needs, and only when he needs it. He suggested a slow build-up approach is needed to determine the traffic density saturation level, and in operation, the overall system needs a function to ensure that level is not reached.

  One pilot pointed out that during the test, pilots were expecting a situation to arise and therefore were prepared for it. In a multitasking operational environment, some form of alert, such as an audio alert, would be required. Both pilots expressed confidence that UML-4 operations could take place under all meteorological conditions except for weather too hazardous for flight. One pilot mentioned a need for the automation to account for all weather hazards in its resolution advisories. The reference prototype automation has such a capability, but the capability was not part of the current testing.

  Neither pilot noticed a difference in runs that varied CD and CR horizons from the default value of three minutes and six minutes respectively. One pilot stated that he was expecting to notice a difference, but he did not. There were no comments regarding feasibility or acceptability for Test Group 6 runs with detection horizons as short as one minute. One pilot commented that he would be comfortable with short time horizons as long as the automation is able to resolve the conflict.

- **Recommendations for next steps**

  The pilots made several suggestions for further research. Judging by responses to many interview questions, both pilots need a refined UI to allow for human/machine interface research to begin. One pilot suggested the display be integrated into the pilot's scan rather than reside on a handheld tablet. One pilot recommended increasing the traffic density in test scenarios. Future tests should also involve equipping the intruder aircraft with FPM automation to investigate dynamic interactions between two FPM-capable aircraft. He also mentioned exploring the current test's assumption that precise along-path trajectory time conformance is required for UML-4 operations.

  When asked, both pilots stated that the use of virtual traffic and the overall test design may have interfered with their ability to make decisions during the test as they would in an operational environment. The research pilots were giving the FPM automation their attention and were not burdened with other tasks. Heavy concentration on the task, such as focusing on the real intruder, was mentioned as a factor. One pilot mentioned creating a more complete environment in a simulator would probably result in different decisions. The other pilot suggested having some conflicts with virtual traffic in test scenarios to give him the feeling of the virtual traffic being a real threat.

One pilot stated that creating "a really bad day" where changes happen frequently, presumably in simulation, would help him build trust in the automation. A pilot also suggested it might be useful to analyze pilot responses in today's Visual Flight Rules operations to determine how far ahead in time they are looking for traffic conflicts.

8.4 Discovery Analysis

This section of the report concerns the Element 3 analysis, discoveries revealed by flight testing. Few really new behaviors were encountered during the flight test itself, however, so these discoveries are supplemented by discoveries made in earlier phases of development of the FPM concept for UAM leading up to the flight test, including the FPM-1 and FPM-2 experiments and the development, simulations, and integration testing performed for IAS-1.

8.4.1 Overall Assessment

Unexpectedly, no new behaviors were discovered during flight testing that had significant effects on the airspace concept. As mentioned in Section 9.1, simulations conducted prior to the test were highly predictive of flight test results.

A guidance system such as that aboard SARA or OPV can easily achieve better conformance to a full 4D flight plan than the FPM concept requires. The focus for future work should be on how tight the conformance really needs to be, and what the trade-offs are between conformance requirements and other factors that determine concept feasibility.

Although the high-density stressor scenarios provided the largest number of traffic aircraft that the airspace model could provide, the traffic density was not great enough in these scenarios to stress the concept or the technology. No reduction in availability of resolution solutions was observed due to high traffic density. No increases in PBGA computing times were observed. To stress the FPM system, the airspace simulation will be required to provide a much higher number of traffic aircraft than the source used for the flight test.

No new or recurring conflicts occurred beyond the CR lookahead horizons used in the test.

Pilots noticed little difference between short lookahead time parameters and the longer lookahead time parameters. They expressed no preference other than they would prefer to have more time to evaluate and select from the resolution options.

8.4.2 Sensitivity to Wind Prediction Errors

Due to the error in handling the predicted wind data (see Section 8.1.2), the flight test unintentionally tested the replanning system's sensitivity to wind prediction errors. In some cases the predicted wind used by AOP underestimated wind speed by more than 20 knots.

For the most part, AOP was robust in the face of wind prediction errors. The main adverse effects of using the wrong wind prediction were that AOP incorrectly predicted the ETA at the RTA point in cases where it would have needed to plan a flight using the minimum CAS in cruise to meet the RTA. In those cases AOP produced spurious RTA resolutions.

It is likely that if the aircraft had had a wider range of airspeeds (for example, if the minimum CAS had not been raised from 60 kt to 75 kt), AOP would have handled erroneous wind predictions even better.

8.4.3 Consistency of Trajectory Prediction

Observations at several stages of preparation for the flight test and in the flight test itself repeatedly illustrated the need for trajectory prediction at each step of flight planning to be consistent with the trajectories planned before and afterward.

This was first observed while developing the UAM-related software for FPM-1 when AOP produced apparently “trivial” lateral resolutions that resolved conflicts with traffic aircraft while making minimal changes in the lateral path of the ownship. This effect was found to be due to different assumptions that the pre-flight planning software and AOP made regarding the distribution of airspeeds among the climb, cruise, and descent segments of the flight plan. This problem was resolved as described in Section 6.4.2.

During the development of the speed-only degree of freedom for PBGA, it was recognized that when planning a CR, the first change in speed must be scheduled at a time after the resolution is expected to be executed, just as the first turn in a lateral resolution or the altitude change in a vertical resolution must be scheduled to occur later in the flight plan. This requirement was addressed by applying the freeze horizon to the aircraft’s speed as well as its horizontal path and its altitude. The flight test, however, pointed out two additional sources of speed changes to which this same principle applies. Sections 8.3.1.2 and 8.3.1.3 describe these cases. There is a need to apply the freeze horizon more generally to speed changes.

8.4.4 Filtering Advisories that Do Not Meet an RTA

When PBGA produced at least one advisory that conformed to the RTA, AOP was designed not to display any nonconforming advisories to the pilot. Due to the mishandling of the wind prediction and the unintended use of six-second RTA conformance in PBGA (see Section 8.3.5), however, many non-conforming advisories were displayed that would likely have been filtered out if the software had been integrated and configured correctly.

The unintended display of non-conforming advisories was a reminder that it is desirable to filter these advisories out of the set presented to the pilot. Alternatively, it would be helpful to indicate to the pilot which advisories do not meet the RTA.

In the future, merely filtering on “in conformance” or “out of conformance” may not be sufficient. In previous interfaces to AOP, the difference between the ETA and the RTA at the constraint point was displayed. Other mechanisms for discarding “useless” advisories may also be considered.

If a future version of AOP is able to relax constraints other than an RTA, the same principles of filtering advisories or informing air crew about the constraint relaxation would apply.

8.4.5 Utility of the Speed-Only Degree of Freedom in PBGA

The usefulness of the speed-only degree of freedom in PBGA appears to depend partly on the type of problem to be resolved. This degree of freedom is designed to enable AOP to plan multiple changes in airspeed during the cruise phase. Resolution trajectories are still required to remain within the bounds of the minimum and maximum airspeeds.

In response to an RTA change, however, the version of AOP used in the flight test was designed to attempt to meet the RTA using a single cruise airspeed if possible before invoking PBGA. In order to cause PBGA to occur, RTA changes in the flight test were designed so that it was impossible to meet the new RTA at the minimum airspeed, the maximum airspeed, or any airspeed in between. This implied that no combination of airspeeds available to the speed-only degree of freedom could ever meet the new RTA. PBGA therefore produced no speed-only resolutions of RTA changes in the flight test.

Speed advisories seem to have limited utility in cases where there is a conflict with an intruder flying the same route as the ownship, since changes in speed alone do not allow aircraft to safely pass on such routes. Speed advisories may enable an aircraft to increase or decrease speed to avoid an overtaking conflict in some cases. Head-on encounters cannot be expected to be resolvable at all.

Despite those restrictions, however, the speed degree of freedom, which had never been implemented prior to work on FPM for UAM scenarios, was surprisingly successful at resolving conflicts in which aircraft cross at an angle.

A disadvantage of the speed advisory as presented to the pilots in the flight test is that it was relatively difficult to understand. Successful use of speed advisories may depend on better representation of the characteristics of an advisory to a pilot.

8.4.6 Utility of the Vertical-Only Degree of Freedom in PBGA

Vertical advisories appear to have a much larger range of usefulness than speed advisories, but still appear to have limited ability to resolve an unmet RTA. In principle, a vertical advisory can change an aircraft's ground speed by changing to an altitude where the winds are different or where there is a significant difference in which true airspeed corresponds to the planned CAS (or Mach number, in the case of higher-altitude aircraft). Since the flight test had a very restricted range of altitudes and had no predicted wind gradients, there were unsurprisingly no successful vertical resolutions of missed RTAs in the flight test.

These considerations raise the question of whether PBGA should even consider speed or vertical resolutions when attempting to resolve an RTA change in UAM operations. Whether vertical resolutions are to be attempted in other environments may depend on the availability of a more detailed wind forecast and the freedom to make large changes in altitude.

8.4.7 Utility of the Hybrid Lateral-Vertical Degree of Freedom in PBGA

In cases where a vertical CR is possible, there is often no optimization to be gained by changing the flight path laterally and the hybrid lateral-vertical advisory tends to be nearly the same as the pure vertical advisory. This phenomenon was noted in the pilots' responses to post-test written questions (Section D.1). In the context of the flight test (short flights, uniform wind field, and conflicts that could be resolved by lateral-only or vertical-only advisories), the hybrid lateral-vertical degree of freedom was not perceived to be useful.

In some other environments, the hybrid lateral-vertical advisory has been found to be desirable. For example, it might be the only possible resolution advisory, or it might be the most efficient resolution advisory. This raises the question of how to determine when hybrid resolutions should be offered.

8.4.8 Additional Maneuver Patterns in PBGA

The original design of the maneuver patterns used by AOP in the flight test did not consider airspace corridors. The observed performance of the flight test system in corridors raised the question of whether a different set of maneuver patterns, designed with the constraints of corridors in mind, might achieve better results. It may also be worth considering ways of combining maneuver patterns from a sequence of invocations of PBGA to better deal with conflicts that are difficult to resolve with a single maneuver pattern.

8.4.9 Lifetimes of Advisories

It is important for AOP to be able to frequently re-evaluate the advisories it has already provided and to cancel any that are no longer suitable to execute, as described in Section 8.3.1.4.

8.4.10 Definition of Fitness of Candidate Advisories in PBGA

In Run M17, a speed-only resolution advisory used what appeared to be extreme airspeeds during the cruise phase. This was determined to be caused by the fact that SARA could not arrive at its scheduled RTA on time without being in conflict with OPV. Because of this, the fitness function in PBGA used the difference between the ETA and RTA for all conflict-free trajectories with no consideration of other optimization criteria. This allowed advisories to include some extreme speed changes as noted in the analysis of MOP 8 in Section 8.2.1

A fitness function that accounted for additional criteria in the case of a missed RTA could be expected to provide a more acceptable resolution in such cases. More generally, the characteristics of fitness functions can be explored to see how well they reflect the actual priorities of aircraft operators.

8.4.11 "Flickering" Traffic Conflicts

In previous environments it was sometimes observed that the predicted time and position at first LOS for a traffic conflict could change significantly when a new CD calculation was performed. In other cases, the prediction of a traffic conflict might disappear altogether, only to reappear a short time later. In these cases it was said that the conflict "flickered."

"Flickering" was considered to be due to the difficulty of making reliable predictions of the 4D trajectories of aircraft that were guided according to three-dimensional flight plans with few or no time constraints. The use of complete 4D flight plans in AOP in the UAM environment might have been expected to solve this problem. Nevertheless, conflicts in test points M17 and M40 "flickered" both in simulation and in the flight test. (Test point M20 had a conflict alert that later disappeared, but this was determined to be due to an incorrect prediction of the ownship's 4D trajectory due to a previously undiscovered defect in BBTG, and this conflict was therefore not considered for analysis.)

The causes of "flickering" in this context seem worth investigating. A possible cause is that AOP recalculated the trajectories that were used for CD too frequently.

8.4.12 Adaptive Lookahead Time

Some of the PBGA results suggested that it would be advantageous for AOP to set a longer or shorter CR lookahead time under certain circumstances. For example, when the ownship is approaching a constraint point that AOP is not permitted to maneuver around, by the time a conflict is detected there may not be enough remaining space for PBGA to plan a successful CR maneuver. In that case it may be preferable for AOP to start probing for conflicts up to the constraint point at some time before the constraint point is within the normal lookahead time.

8.4.13 Load-Balancing for PBGA

As observed in Section 8.3.4.3, there is an opportunity for better load-balancing of the computations performed by PBGA. In addition, close examination of the records made by PBGA indicated that threads started in one CR cycle were not always terminated when the next CR cycle began, although their results were always discarded.

8.4.14 Assumptions Inherited from Previous Applications of AOP

Some of the adverse effects observed can be attributed to features developed for different operating environments. For example, the handling of an RTA change in AOP was originally developed in an environment in which a flight management system (FMS), not AOP, produced the flight plan to be flown by the guidance systems. In this previous environment, by the time AOP received the RTA, the FMS had already changed the aircraft's speed to conform to the RTA. The concerns about AOP's handling of RTAs during the flight test could not easily have been anticipated in that environment.

8.4.15 Conceptual Concerns

The FPM system as implemented for the flight test was designed to address critical concerns of the FPM concept. Not all concerns could be addressed in the scope of this work. Reconsideration of at least two aspects of AOP's flight planning algorithms is recommended.

8.4.15.1 Relaxation of 4D Conformance

As mentioned previously, the 4D trajectory conformance tolerances used in the flight test appear to be tighter than necessary. In addition to loosening the tolerances in general, specific parts of the flight plan, such as climbs and descents, may benefit from having larger TPUBs than other parts of the flight plan.

Simulation results suggest that 4D conformance tolerances could be greatly increased without sacrificing the airspace density used in the flight test. For example, looser TPUBs in the vertical and along-path dimensions during climb and descent could enable the aircraft's guidance systems to follow the most efficient flight path for the conditions that are encountered.

8.4.15.2 More Frequent Replanning

The once-and-done nature of AOP's flight planning as implemented for the FPM simulations and for the flight test — recomputing the flight plan only in response to specific events such as a conflict or an RTA change — worked well in the flight test but may be inadequate for longer flights. Further work on the implementation of the FPM concept is required in order for AOP to compute and broadcast new flight plans at appropriate intervals of time to adapt to changing conditions or to differences between observed and predicted conditions. But the strategic nature of the FPM concept requires that AOP not replan too frequently. Previous work on flight path optimization with AOP (Ref. 24) is relevant.

8.4.16 Challenges in Describing the Concept

Integration efforts for the flight test revealed ways in which the language used to describe the FPM implementation for UAM could lead to misconceptions.

The use of the term "4D flight plan" may have undesirable connotations. AOP is designed to choose optimal airspeeds in each phase of flight, taking into account various factors that affect aircraft performance, such as aircraft gross weight and predicted winds and air temperatures. The use of ground speed in the 4D flight plans shared with other parties in the airspace was intended only to enable other aircraft to plan safe trajectories without sharing a much larger set of data used in AOP's flight planning, but on many occasions collaborators unfamiliar with AOP's internal functions assumed that AOP ignored many of these considerations.

It is important to communicate that while the flight plans shared with other aircraft (using EUTL as the trajectory description in this implementation) imply a 4D path described by ground speeds, this is done mainly to avoid the complications of using airspeed to predict traffic positions. There is no application in which AOP would plan to fly at a constant ground speed in any phase of flight. Instead, in the flight test AOP attempted to hold CAS constant, resulting in a lower ground speed when the aircraft turned into a headwind and a higher ground speed when the headwind was less. The climb and descent phases of flight were assumed to require lower airspeeds than the cruise phase.

Misconceptions about AOP's flight planning criteria led to difficulty explaining the information that AOP actually needed to perform its functions correctly, and may contribute to a false sense that a system such as AOP can never create practical flight plans.

8.5 Simulation Validation Analysis

As discussed in Section 2, the primary objective of the simulation validation was to confirm that the simulation environment reflects the real world. Real-world data is typically difficult, time consuming, and expensive to collect. Confirming that the simulation environment is reflective of real-world situations and performances enables exponential scaling of data collections, generating mass data in a timely and affordable fashion. The validation consisted of three comparisons:

1. Replicating flight test data in simulation (Section 8.5.2).
2. Comparing wind prediction conditions in simulation (Section 8.5.3).
3. Comparing the effects of resources and repeatability in simulation (Section 8.5.4).

Each comparison served a separate purpose to establish whether the simulation environment is providing quality data. The Flight Test Comparison attempts to replicate exactly what happened during the flight test. The researchers noticed several surprising behaviors including multiple conflicts in a single run, conflicts with virtual background traffic, and flickering of the conflicts. To reproduce exactly what happened during the flight test, there was a need to replicate the incorrect wind input that went unnoticed during the flight test (as mentioned in Section 8.1.2). The Wind Input Comparison section involves comparing the simulation runs that replicated the flight test (incorrect wind input or Wind Error) to simulation runs that corrected the wind input error (Correct Winds). The final section, Repeatability and Resources Comparison, is meant to determine the reproducibility of the simulation results. Specifically, whether repeating the same simulation sequentially produces identical results, and whether modifying the computational resources affects the simulation behavior.

8.5.1 Methods

The simulation environment used for the validation was a NASA hosted Amazon Web Services platform that enabled switching between different virtual machines for running the software. Here, the primary machine was an x64 Windows based PC with an Intel Xeon Platinum processor with 16 cores and 32 GB of RAM (referred to as "High Resource" machine). The flight-test had a modular setup with TCP sockets connecting the three primary pieces of software relevant for the FPM research (AOP, FPM GUI, and auSimAPI). For the flight test, each of these pieces of software was running on a different machine. In the simulation environment, this software was running on the same machine and utilizing shared memory instead of sockets.

Table 12. Simulation Validation Maneuver List. This table lists the maneuvers run as part of the simulation validation in sequential order from top to bottom. The Maneuver ID corresponds to the same maneuvers laid out in Table 6. The Resolution Execution column states whether that run had an executed resolution. The Wind Mode column states whether the simulation either replicated the wind prediction error (Wind Error), had the correct wind setup (Correct), or it did not matter because there were zero winds (Zero Winds). The Wind File column tells us which wind file was used for each run. The simulations used the same wind files used during the actual flight test for the respective maneuvers. The Machine Resources column states whether the maneuver was run in the 16 core 32 GB RAM environment (High CPU) or the 4 core 8 GB RAM environment (Low CPU).

Maneuver ID	Resolution Execution	Wind Mode	Wind File	Machine Resources
M10a	No Execute	Wind Error	10Kt-F300	High CPU
M17	Execute	Correct	10Kt-F180	High CPU
M1	Execute	Wind Error	20Kt-F240	High CPU
M4	Execute	Zero Winds	00Kt-F000	High CPU
M1	No Execute	Wind Error	00Kt-F000	High CPU
M5	No Execute	Zero Winds	00Kt-F000	High CPU
M8	Execute	Wind Error	10Kt-F180	High CPU
M13a	No Execute	Wind Error	10Kt-F300	High CPU
M3	Execute	Zero Winds	00Kt-F000	High CPU
M40	Execute	Wind Error	10Kt-F180	High CPU
M42	Execute	Wind Error	20Kt-F240	High CPU
M19	Execute	Wind Error	10Kt-F240	High CPU
M21	Execute	Correct	20Kt-F240	High CPU
M32	Execute	Wind Error	10Kt-F240	High CPU
M7	No Execute	Wind Error	10Kt-F180	High CPU
M17	Execute	Wind Error	10Kt-F180	High CPU
M21	Execute	Wind Error	10Kt-F180	High CPU
M3	Execute	Wind Error	20Kt-F240	High CPU
M13a	Execute	Correct	10Kt-F300	High CPU
M1	Execute	Correct	20Kt-F240	High CPU
M10a	No Execute	Correct	10Kt-F300	High CPU
M39	Execute	Correct	10Kt-F180	High CPU
M42	Execute	Correct	20Kt-F240	High CPU
M39	Execute	Wind Error	10Kt-F180	High CPU
M15	Execute	Wind Error	10Kt-F180	High CPU
M5	No Execute	Wind Error	00Kt-F000	Low CPU
M1	No Execute	Wind Error	00Kt-F000	Low CPU
M10a	No Execute	Wind Error	10Kt-F300	Low CPU
M1	No Execute	Wind Error	00Kt-F000	High CPU
M1	No Execute	Wind Error	00Kt-F000	High CPU
M1	No Execute	Wind Error	00Kt-F000	High CPU
M1	No Execute	Wind Error	00Kt-F000	High CPU
M1	No Execute	Wind Error	00Kt-F000	High CPU

Another primary component of replicating what happened in flight is replicating the behavior of the pilot. This involved executing presented resolutions at a specific point in time. To enable this, the researchers created a UI that grabbed information from a meta-data file that provided information about the runs completed during the flight test. Specifically, PBGA execution time, PBGA resolution type, RTA Speed Change execution time, execution cycle, and the wind file were pulled from the meta-data file for the selected run and presented this information to the researcher. The UI was programmed to automatically countdown, and triggered off a background process that corresponded to T_0 of research mode, so that the simulation resolution execution could be synchronized with flight test data. Resolutions were executed from the execution cycle shown in the flight test data whenever possible.

Prior to running a maneuver, the researchers would check and modify the wind parameters, if necessary, to achieve the desired wind behavior in the simulations. For example, if the goal was to replicate the flight test incorrect wind input, the researcher would go into the wind file that corresponded to the wind conditions on the day of flight test for the respective maneuver and set the predicted wind info to zero. This effectively prevented AOP from taking the wind into account when performing its functions.

Due to the time-consuming nature of reproducing exactly what happened in the flight test in the simulation environment, a subset of the data collected in the flight test was selected for simulation. See Table 12 for the complete list of maneuvers run in the simulation validation. The table shows the maneuvers run in sequential order from top to bottom.

8.5.2 Flight Test Comparison

This section presents an overview of the comparison between flight test and the simulation data. The researchers completed 17 runs for this comparison. The runs were loosely arbitrarily chosen. Some runs were chosen based on some odd observations that were made. For example, maneuver M10a had a conflict with a virtual traffic aircraft, and maneuvers M21 and M42 had resolutions occur before the conflict. The hypothesis here was that there would be no difference between flight and simulation data. As stated earlier, the dataset required reproduction of the incorrect wind input in the simulation, as well as the replication of the pilot input for executed resolutions.

Figure 19 shows the results of the 17 maneuvers simulated for comparison with the flight test data. The y-axis lists those 17 runs with the wind file that was used during the flight test. Within each maneuver's row, the flight and simulation resolutions and conflicts are displayed over time. In general, the simulation largely reproduced the flight test data. Most of the runs had identical outcomes in regard to the number of resolution cycles, types of resolutions, and the timing of the resolutions and conflicts.

There were a few occasions where the simulation data diverged from the flight test data. The most obvious case of this is maneuver M21. During the simulation, AOP did not compute resolutions as frequently after resolving the conflict as it did during the flight test. It should be noted that M21 was a maneuver that was heavily affected by the incorrect wind input (as described in Section F.4.1), that the "missing" resolutions in simulation were all RTA resolutions, and that even in the flight test these resolutions were less frequent than the unplanned RTA resolutions in some other runs. It was noted that in the flight test, the active flight plan in M21 was only barely out of conformance after the CR according to the logic that triggered PBGA. Maneuvers M15, M32, and M40 had slightly offset conflict onsets with respect to the same maneuvers in flight. Maneuver M5 had an additional conflict with a virtual traffic aircraft during the flight test, but in simulation the conflict with OPV continued throughout.

The researchers chose a subset of metrics to quantify on the simulation data to confirm that the simulation is providing comparable data that are within acceptable thresholds. All simulated runs analyzed

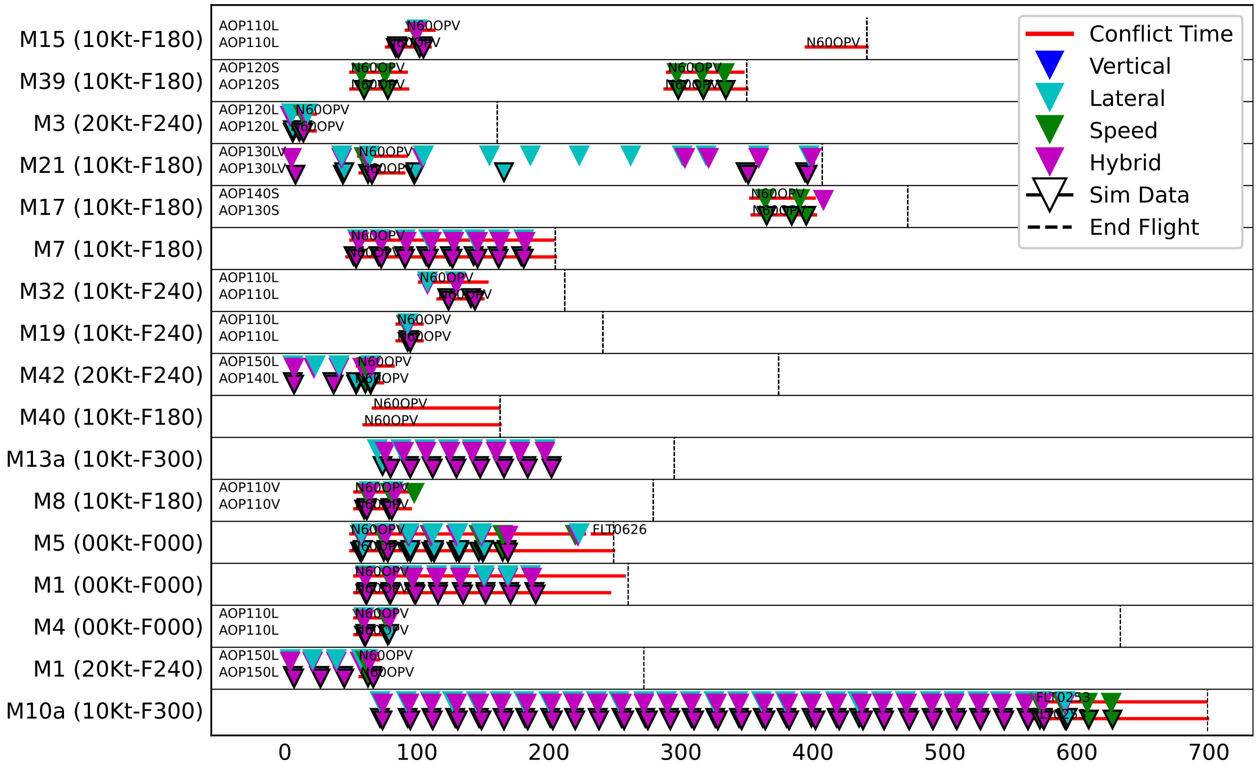


Figure 19. Overview of flight vs. simulation data that reproduced the wind conditions in the flight test. The y-axis indicates the runs that were run in both simulation and flight, while reproducing the wind conditions observed in flight (wind condition in parentheses). The first text within each run's row indicates which resolution was executed, if any. The number corresponds to the cycle the resolution was executed from (i.e. 110=1st cycle). The following letters correspond to the type of resolution that was executed (L=Lateral; V=Vertical; LV=Hybrid; S=Speed). The occasions where it is blank indicates no resolution execution. The flight data is always above the simulation data for a respective run, and the simulation resolutions have black outlines to further distinguish the two. The vertical dashed line indicates the end of the flight data. The conflict is indicated by the red line, while the text at the beginning of the red line indicates the call sign of the aircraft causing the conflict.

were replicated with incorrect wind input. The results from the subset of metrics included:

- Metric 1: False Detection Rate
  No change between flight and analyzed simulated runs. M10a was the only run to show a false detected conflict (M10a, M1, M4, M8, M42, M19, M32, M17, M21, M3, M39, M15).

- Metric 3: Actual vs. Predicted CPA Error
  In seven analyzed runs, there is never perfect replication of the CPA, in either actual or predicted. Figure 20 shows the comparison between the simulation maneuvers and the matching flight test maneuvers. There does not appear to be a trend in either direction (for example, a lower CPA in simulation than in flight).

- Metric 8: Rate of First CR Cycle Resolution for all conflicts
  No change between flight and analyzed simulated runs (M1, M4, M8, M15, M42, M19, M32, M17, M21, M3, M39).

- Metric 11: Conflict Prevention Rate

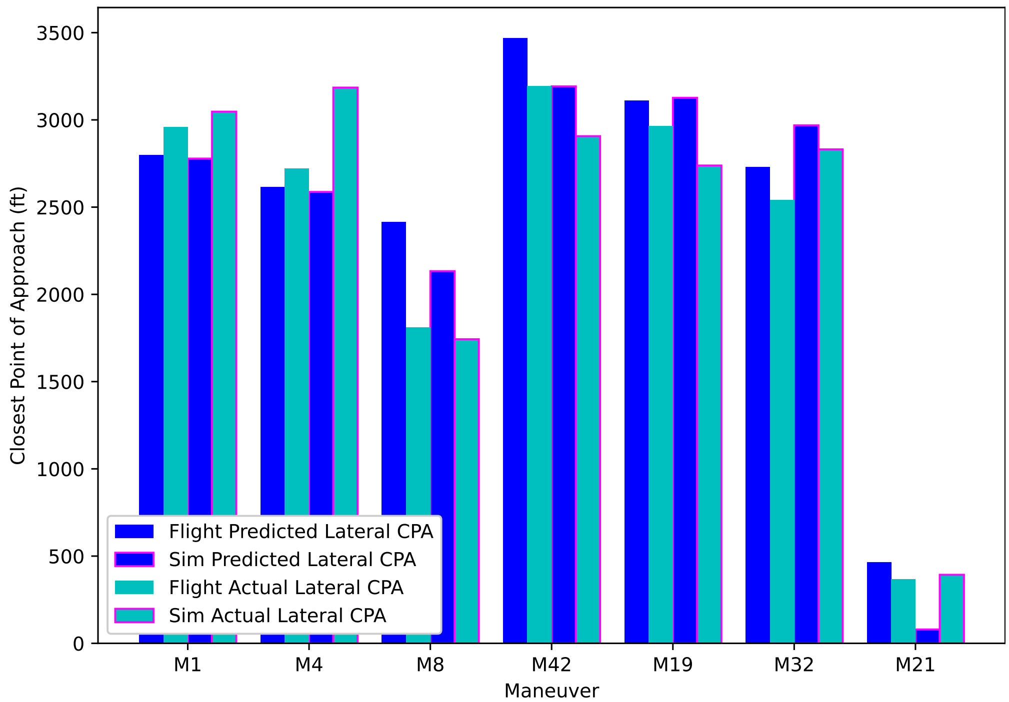


Figure 20. Simulation vs. Flight Test CPA Error comparison. Only lateral CPA is shown because this was the source of the most error between predicted and actual CPA. The magenta highlighted bars correspond to the simulation data (the two right-most bars). The blue bars are the predicted CPA. The cyan bars are the actual CPA.

In the simulated runs analyzed all runs prevented conflicts in their range of validity, similar to flight test.

- Metric 15: CR Performance Evaluation
  There is a quantitative difference in CR computation completion time between flight test and simulation data, shown in Figure 21. The mean and the variability are slightly increased in simulation data, where flight data took 5.4 ± 2.01 s and simulation data 6.7 ± 3.09 s.

A primary objective was to quantify how similar the simulation was to the flight test results. As mentioned in Section 7.1, the batch simulations prior to flight involved a high-level check of the simulation results. Additionally, the metrics described above were not developed with simulation and flight test comparisons in mind, and therefore do not provide complete insight into the similarities and differences observed in Figure 19. Therefore, there was a need to develop a separate, custom metric to describe the simulation similarity to flight test data, a figure of merit. This figure of merit is composed of four separate parameters that each describe an aspect of the results shown in Figure 19. The four parameters include conflict onset similarity, total cycle similarity, resolution type similarity, and resolution timing similarity.

Equation 1 defines $conflict\_onset\_similarity$, where it is the percent similarity between flight and simulation conflict onset times ($conflict\_onset_{flight}$ and $conflict\_onset_{sim}$, respectively), with respect to the cycle time. The CR cycle time, denoted $cycle\_time$, was 20 seconds in most maneuvers completed during flight test and 15 seconds for maneuvers M19 to M30, inclusive.

$$conflict\_onset\_similarity = 1 - \frac{\left| conflict\_onset_{flight} - conflict\_onset_{sim} \right|}{cycle\_time}. \quad (1)$$

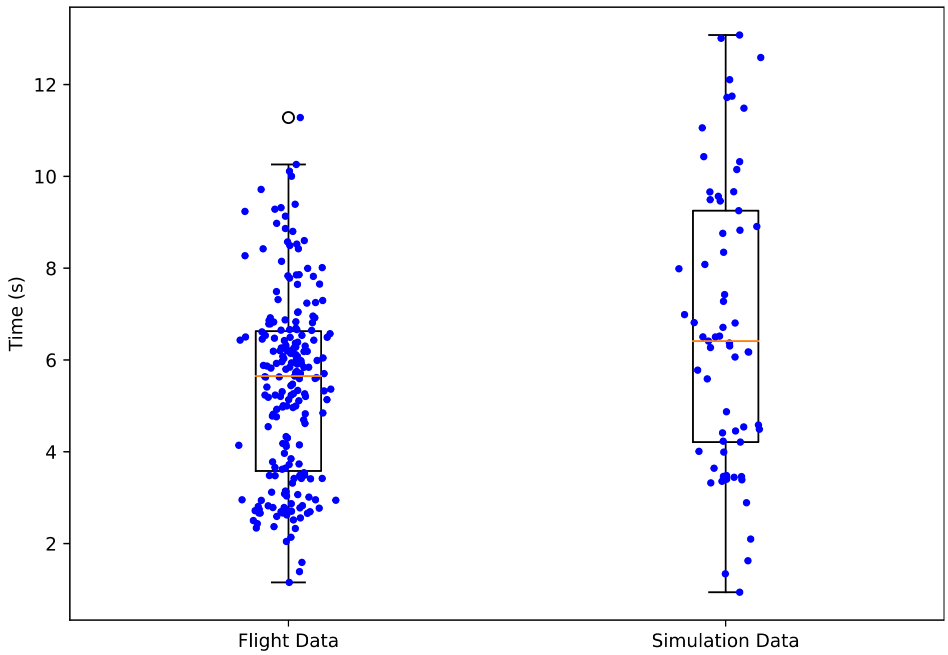


Figure 21. Display of how much time it took for a CR cycle to complete for every CR computation completion in flight test and simulation data.

Equation 2 defines $total\_cycle\_similarity$, where it is the percent similarity between flight and simulation cycle counts (number of resolution cycles in the maneuver), with respect to the flight cycle count.

$$total\_cycle\_similarity = 1 - \frac{\left|cycle\_count_{flight} - cycle\_count_{sim}\right|}{cycle\_count_{flight}}. \tag{2}$$

Equation 3 defines $resolution\_type\_similarity$, where it is the percent similarity between the flight and simulation resolution degrees of freedom (res_vector_{flight} and res_vector_{sim}, respectively), averaged across every resolution cycle of flight ($cycles\_flight$). Each resolution vector (res_vector) was a binary 1×4 vector representing whether the degree of freedom was present in the order of lateral, vertical, hybrid, and speed resolution types. The denominator here is constant 4, as there were only ever 4 possible degrees of freedom for resolutions.

$$\begin{aligned} resolution\_type\_similarity &= \frac{\sum_{n=1}^{cycles\_flight} cycle\_similarity(res\_vector_{flight}[n], res\_vector_{sim}[n])}{cycles\_flight} \\ \text{where} \quad &\begin{cases} cycle\_similarity(u, v) = \frac{1}{4}\sum_{i=1}^{4}[u_i = v_i], \\ [P] = \begin{cases} 1 & \text{if } P \text{ is true,} \\ 0 & \text{if } P \text{ is false} \end{cases} \quad \text{(definition of the Iverson bracket).} \end{cases} \end{aligned} \tag{3}$$

Finally, Equation 4 defines $resolution\_timing\_similarity$, after the resolution time differences from the respective (flight vs. sim) conflict onset time were calculated, denoted res_diff. The resolution timing score was based on the percent similarity, with respect to the cycle time, again denoted $cycle\_time$. The percentages were averaged over all compared resolutions. Resolution timing scores were only included if both the flight and simulation contained the same resolution type in the same resolution cycle.

$$resolution\_timing\_similarity = \frac{\displaystyle\sum_{n=1}^{resolutions\_flight} \left(1 - \frac{\left|res\_diff_{flight}[n] - res\_diff_{sim}[n]\right|}{cycle\_time}\right)}{resolutions\_flight}. \tag{4}$$

All similarity parameters were then averaged across maneuvers. Table 13 displays the results for each parameter and the total similarity score. These maneuver labels correspond to the same maneuvers in Figure 19, but M21, M40, and M13a are excluded. Maneuvers M40 and M13a are excluded due to the lack of resolutions and conflicts, respectively, and M21 is excluded because the resolution cycles were obviously so different. The average total similarity score was 89.6% (averaged across all runs).

It must be reiterated that this was a cursory attempt at simulation validation, and this also applies to the figure of merit. There are some limitations with its implementation, for example the Cycle Onset Similarity is biased by how many cycles there are in a particular maneuver. M15 only had one cycle in flight, and two in simulation, therefore the similarity score was 50%. Additionally, the Total Cycle Similarity is sensitive to error in timing the resolution execution. Again, this most obviously affected M15. The researchers prioritized executing the resolution at the same time past T_0 as was done in flight test, and because the conflict appeared earlier in simulation, the same time past T_0 allowed another resolution cycle to appear. These, among other components, could be improved upon in future attempts at simulation validation.

Table 13. Results of the figure of merit that was used to quantify the similarity between simulation and flight test results. The scores are represented in decimal form with 0.0 representing 0% and 1.0 representing 100%. The Total Similarity Score is the average of the four parameters (Conflict Onset Similarity, Total Cycle Similarity, Resolution Type Similarity and Resolution Timing Similarity).

Maneuver ID	Conflict Onset Similarity	Total Cycle Similarity	Res Type Similarity	Res Timing Similarity	Total Similarity Score
M15	0.20	0.50	0.63	0.98	0.5875
M39	1.00	1.00	0.96	0.93	0.9736
M3	0.80	1.00	1.00	0.94	0.9354
M17	0.95	1.00	0.96	1.00	0.9766
M7	0.85	1.00	1.00	0.91	0.9408
M32	0.30	1.00	1.00	0.94	0.8094
M19	1.00	1.00	1.00	0.92	0.9791
M42	0.90	0.60	0.79	0.50	0.7146
M8	1.00	1.00	0.92	0.90	0.9556
M5	1.00	0.88	0.92	0.96	0.9393
M1	1.00	1.00	1.00	0.94	0.9845
M4	1.00	1.00	0.92	0.95	0.9667
M1	0.95	0.80	0.90	0.87	0.8788
M10a	0.85	1.00	1.00	0.78	0.9084

Despite the undesirable affects of the flight test's incorrect wind input, the simulation did a good job of replicating the flight test data ($\sim$90% similar). There were some instances where the simulation did not align identically with the flight test data, but there could be several explanations for this. The differences in Metric

3 between flight test and simulation data potentially point to aircraft trajectory conformance being different between flight test and simulation. Not only are the flight dynamics different, but the guidance systems are also different, potentially contributing to some of the differences. This is not something specifically analyzed here, but should be taken into consideration for future analyses. Some other potential explanations revolve around software architecture and hardware resources. The resource aspect is delved into in a following section.

8.5.3 Wind Input (Wind Error vs. Correct Wind) Comparison

One of the preliminary explanations for the odd behaviors observed in the flight test data was that it was a result of the incorrect wind input. This section serves to confirm whether that is true. The researchers compared six maneuvers that were simulated for the flight test comparison with the incorrect wind input, to the same maneuvers that were simulated with the correct wind input. Maneuvers were chosen that had RTA conformance issues (presumably due to the incorrect wind input), and a few others that at least had some winds (more than 0 kt).

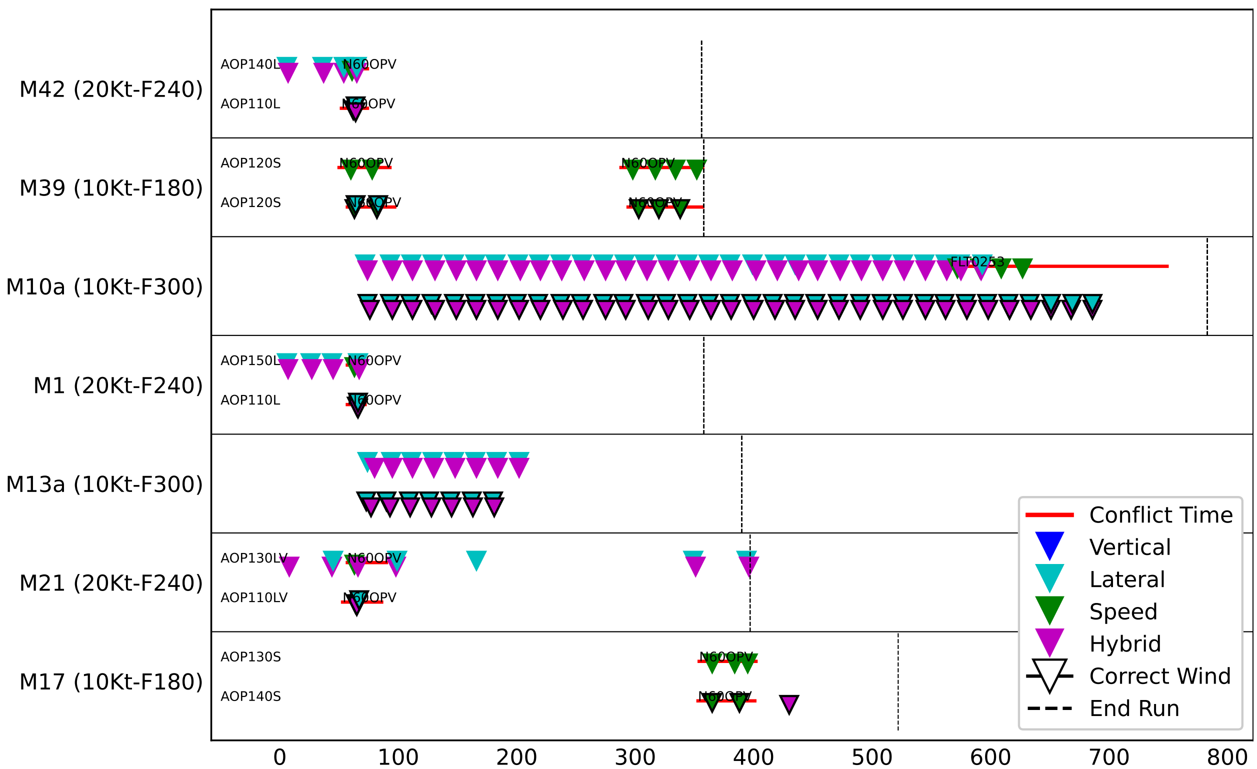


Figure 22. Overview of the comparison between Wind Error vs. Correct Wind data from simulation. The y-axis indicates the runs that were completed in both Wind Error and Correct Wind conditions. The first text within each run's row indicates which resolution was executed, if any. The number corresponds to the cycle the resolution was executed from (i.e. 110=1st cycle). The following letters correspond to the type of resolution that was executed (L=Lateral; V=Vertical; LV=Hybrid; S=Speed). The occasions where it is blank indicates no resolution execution. The Wind Error is always above the Correct Wind for a respective run, and the Wind Corrected resolutions have black outlines to further distinguish the two. The vertical dashed line indicates the end of the flight data. The conflict is indicated by the red line, while the text at the beginning of the red line indicates the call sign of the aircraft causing the conflict.

Figure 22 displays the results in a similar format to the previous section's Flight Simulation comparison figure. The y-axis displays the maneuver and the wind conditions in parentheses, and within each maneuver row the simulated Wind Error (on top) and Correct Wind (on bottom) resolutions and conflicts are shown. In three cases (M42, M1, M21) the Correct Wind made a substantial difference in the resolutions presenting outside of the conflict notification. This implies that the incorrect wind input was indeed responsible for the RTA conformance in all of the maneuvers that exhibited this behavior. Furthermore, in M10a, the conflict with the virtual traffic does not occur in the correct wind simulation. The other maneuvers were not expected to show any change from incorrect wind input to correct wind, and that proved to be true.

This analysis proved that a number of the odd behaviors observed during flight test were corrected when running the simulation with the correct wind setup. The Correct Wind results look more sensible, and indicate that flight test results might benefit from consideration of the Correct Wind simulation results.

8.5.4 Repeatability and Resources Comparison

This section is meant to get a sense of how repeatable AOP behavior is, as well as whether the computational resources affect the outcome. For the resource comparison the researchers switched the virtual machine to a setup that used the same Intel Xeon Processor, but with four cores instead of 16, and 8 GB of RAM instead of 32 (referred to as "Low Resource"). For the resources test, the researchers chose three maneuvers that contained many resolutions, all of which involved no CR execution. For the repeatability test, the researchers chose a single maneuver that had no wind, produced several resolutions, and no execution.

Figure 23 displays the three maneuvers that were simulated in both High and Low resource environments. The High Resource run is shown on the top, and the Low Resource Run is shown on the bottom. Interestingly, there is virtually no difference in conflict time, in all three maneuvers, but the resolutions are delayed substantially in all three. Moreover, in M1 and M5 the type of resolution and the timing within a resolution cycle is much more variable than in the High Resource environment.

Figure 24 displays five different completions of the same run, completed sequentially within about one hour of time. The bottom run was completed first and the top completed last. The results do show slight differences over the five runs. The first three runs look very similar, in both CD and CR timing, but the fourth and fifth diverge from the first three with an earlier CD and CR generation.

Figure 25 displays the CR cycle completion time for the flight test data, the High Resource, and Low Resource simulation data. There is a striking increase in completion time for the low resource data. The variability also increased, but it should be reiterated that this is a small sample size and just a cursory simulation validation. For comparison the flight data took 5.4 ± 2.01 s and high resource simulation data 6.7 ± 3.09 s, while the low CPU CR cycles took 31.9 ± 11.1 s.

Resources clearly impact the behavior and outcomes of AOP. This may contribute to some of the differences that were observed in the original simulation and flight test data comparison. A way to confirm this would be to use the same hardware used in the flight test, which can be done by running the simulations in the Systems Integration Lab. Interestingly, regardless of whether the resources are substantially different, repeating the same maneuver does not yield identical results every time. Given static resources, the behavior is similar across runs, but the timing may not be consistent. This could be related to running on a Windows platform that makes it difficult to control background processes, but this will require further investigation.

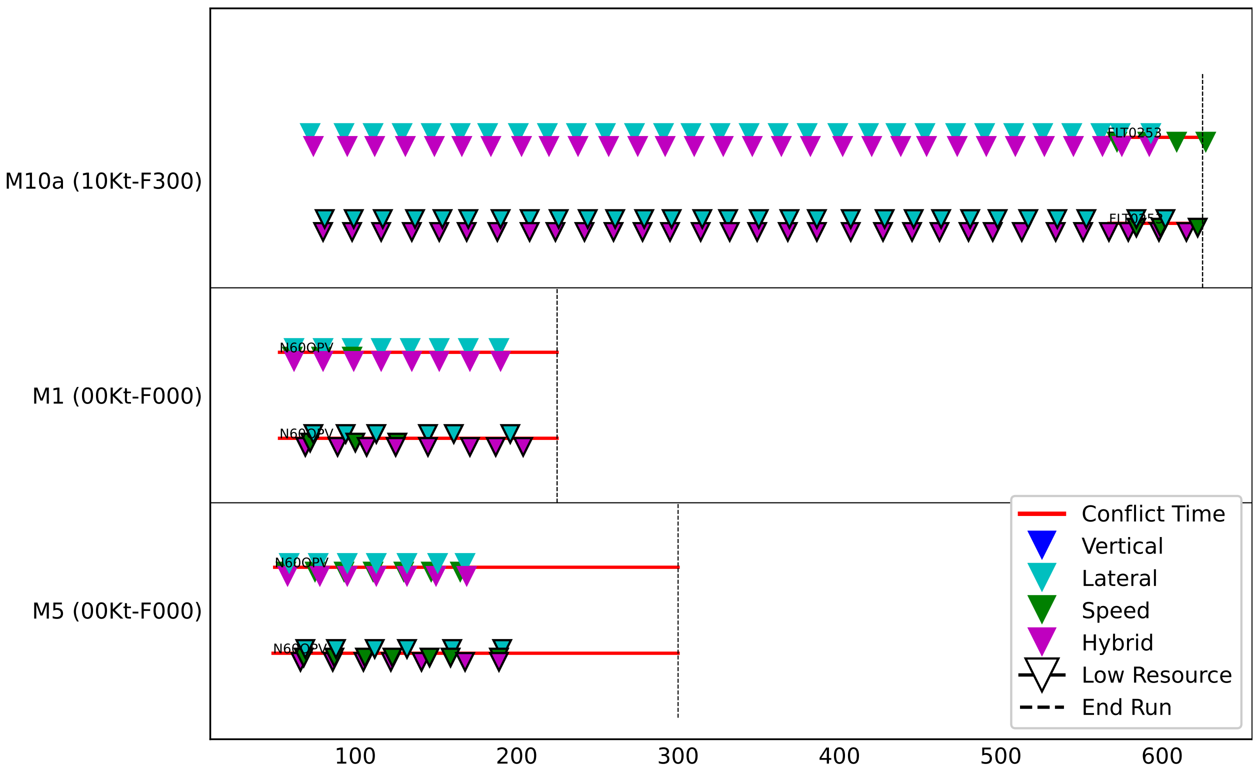


Figure 23. Overview of simulation data that compared different computational resources within the simulation environment. The y-axis indicates the runs that were completed with High Resource and Low Resource environments, while reproducing the wind conditions observed in flight (wind condition in parentheses). High Resource data is always above, and the Low Resource resolutions have black outlines to further distinguish the two. The vertical dashed line indicates the end of the flight data. The conflict is indicated by the red line, while the text at the beginning of the red line indicates the call sign of the aircraft causing the conflict.

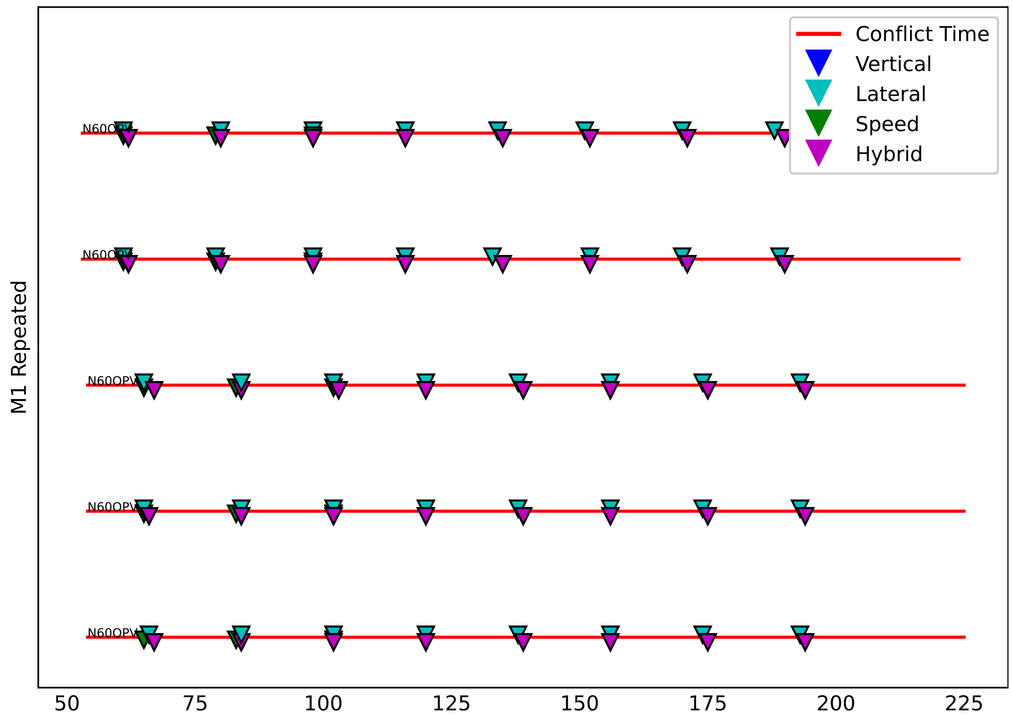


Figure 24. Overview of simulation data that compared the same run repeated five times within one hour. The conflict is indicated by the red line, while the text at the beginning of the red line indicates the call sign of the aircraft causing the conflict.

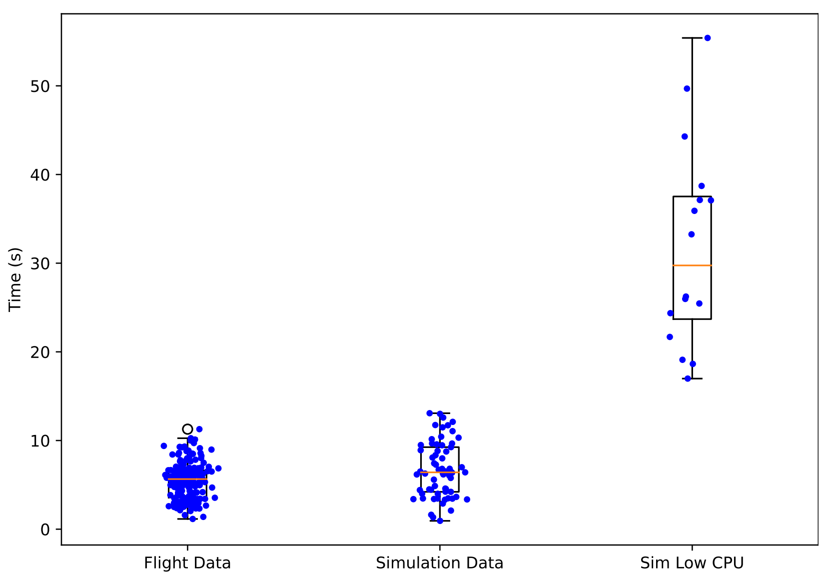


Figure 25. Display of how much time it took for a CR computation to complete for every CR in flight test, high resource simulation data, and low resource simulation data. The Low CPU data only contains data from two maneuvers (M1 and M5 from Figure 23).

9 Conclusions

This activity achieved the first successful flight test of airborne automation that performs strategic flight path deconfliction and in-flight planning to support traffic management. The test produced an initial verification of the automation's functions. It generated insights regarding Flight Path Management (FPM) concept feasibility for future high traffic density operations. It also revealed needs for technology refinement and continued research. Feasibility and readiness conclusions are presented below, followed by recommendations for next steps.

9.1 Insights and Considerations

Major insights gained from the activity and considerations for interpretation of results are summarized below.

9.1.1 Technology Functional and Computational Performance

- The strategic conflict detection (CD), conflict resolution (CR), and conflict prevention (CP) functions performed well in flight in the near-open airspace conditions for which they were designed. All success criteria defined for the test were met. Resolution advisories were always flyable.
- Strategic resolutions complied with assigned required time of arrivals (RTAs) reliably. RTA change assignments up to one minute, the maximum tested, were also handled by the automation up to approximately 300 seconds prior to the destination.
- Traffic densities tested in flight and simulation were not high enough to reveal any traffic density limits to the concept or the enabling FPM technology.
- No future breakthroughs in computational power will likely be required. The prototype technology easily handled 330 traffic aircraft operating on a consumer-grade four-core processor. There appears to be no need to filter traffic to increase computing speed.

9.1.2 FPM UAM Maturity Level 4 (UML-4) Concept

- The flight test produced no new concept behavior discoveries. Simulations conducted prior to the test were highly predictive of flight test results.
- The small separation minima assumed for the future UML-4 environment caused no issue, given the tight trajectory conformance used in the test.
- Pilot feedback suggests human operators can be in or on the FPM loop at UML-4. Pilots had at least 15 seconds to evaluate and execute the automation's advisories for all time horizon parameter settings except the short settings.
- Autonomous Operations Planner (AOP) never presented an advisory that pilots were uncomfortable with or felt was unsafe.
- While pilots were able to successfully integrate into the FPM concept at UML-4, pilots expressed interest in having more time to allow for a more complete evaluation of each advisory.

- The pilots expressed interest in being informed on which AOP resolution is most efficient (based on optimization criteria) within a given set of advisories.
- The pilots felt that some path changes AOP produced were not intuitive or natural for a pilot to make.
- The baseline, long-, and very-long-duration time horizon parameters tested revealed no issues in functional performance and no pilot comments regarding time preference.
- Traffic intent changes revealed no issues for changes as close as one minute prior to first loss of separation (LOS), the minimum tested.
- The short time horizons tested provided very short, potentially unacceptable, resolution display time durations in a few cases.
- FPM functions will require substantial refinement for operations within airspace corridors. Test results are inconclusive, but a different approach may be required for high-density corridor operations.
- The ground-referenced a four-dimensional (4D) aircraft guidance assumed for the concept and implemented for the flight test produced highly successful performance for all tested FPM functions. Satisfactory results may be possible without a requirement for a 4D guidance.

9.1.3 Technology Prototype

- The flight test revealed several software defects and many opportunities for technology refinement.
- The technology's constraint relaxation capability successfully optimized the level of relaxation for the constraint, but it did not consider other optimization criteria when doing so.
- The FPM research prototype's architecture for the allocation of computing resources to CR and RTA resolution could be refined.

9.1.4 Simulation Validation

- While results were not identical, a multi-aircraft FPM traffic simulation reproduced concept behaviors observed in flight.
- Validation results suggest several observed issues during the flight test were caused by the wind input error discovered post-flight.
- The simulation used in the validation exercise produces outputs that are highly sensitive to computing power of host platforms.

9.1.5 Flight Testing

- Development and integration of real and virtual aircraft, future airspace models, scenario generation capabilities, and test facilities required labor-intensive iterative preparation.
- Designing and executing test maneuvers that reliably created live conflicts proved to be a challenge.
  - Identical trajectory generators in design and execution phases were required to ensure conflicts happened and to prevent trivial encounter solutions

  – Accounting for daily variable winds required preparing an extensive matrix of winds cases

- Overtake test cases approached geographic limits of the test range. A larger test range would increase testing ability, especially in high winds.
- Future tests may require a data link fully dedicated to the system under test

9.2 Technology Maturity Assessment

All functional performance success criteria for maneuvers in the open airspace environment were met or exceeded. The research prototype automation technology reliably supported the core functions of intent-based CD, strategic CR, CP, and arrival time compliance. The technology-equipped test aircraft successfully maintained separation with all traffic. It also kept clear of virtual restricted airspace regions, which made up over 11% of the airspace environment. The FPM technology generated resolution trajectories that were always flyable within procedural and performance limits of the hosting aircraft. Through an engineering user interface (UI) that displayed resolution advisories and no other decision assistance, the automation provided pilots enough time to evaluate and execute resolutions.

Test maneuvers involving operations within flow corridors were less successful. Results suggest a specific trajectory management approach and its enabling automation technology need to be developed to support high-density operations within corridors. The approach may involve aircraft in-trail spacing additions to the current capabilities. At a minimum, further simulation research and technology refinement is necessary to establish feasibility of the current approach, which was designed to support arrival metering.

Supporting simulation capabilities are a critical element of the complete toolset needed for continued advancement. Simulation-based research has a goal to explore the interactions of hundreds of FPM-capable aircraft, studying system behaviors for hundreds of operational scenarios while varying dozens of design parameters. Toward this goal, the initial validation performed for this study made use of a cloud-based traffic simulation that can scale up cost-effectively. The validation indicates such a simulation can reproduce general behaviors observed in flight. It also showed that the current research prototype automation's outputs are highly impacted by the computing capabilities of its hosting platform. To duplicate flight results precisely, the automation must be hosted by a simulation platform having a capability very similar to the flight platform. Other simulations operated by NASA can host the flight hardware. Such simulations should produce results that closely duplicate flight results.

The initial validation's results suggest that with appropriate control of host platform capabilities, the current cloud-based simulation can be used to identify behavioral trends of airborne system elements at both a vehicle level and an airspace system level. Enhancements in simulation sophistication will be needed to conduct system requirements research. A determination of separation minima will require increases in aircraft model and flight guidance sophistication. A full understanding of information sharing will require appropriate representations of communications systems, latencies, and bandwidth availability. A complete understanding of overall system behavior, including an understanding of distributed responsibility and authority between the airborne and ground-based system elements, will require inclusion of the ground-based resource management capabilities.

Within the limits of this study, the technology has achieved a level of maturity equivalent to Technology Readiness Level (TRL) 5. Limitations include:

- As previously explained, successful results are limited to operations in open airspace containing a low fraction of the total area that is not available for use.

- Only one aircraft was equipped with FPM technology in the test. The interactions between multiple self-managed aircraft were not studied. Right-of-way rules that determine which aircraft has priority were not explored, nor were pilot or system behaviors resulting from use of priority rules.
- A 4D ground-referenced flight guidance was used in the test. System behaviors and intent-sharing impacts resulting from relaxed flight guidance requirements were not explored.
- The technical challenges and certification requirements associated with collaborative and responsible automation are not considered.
- Data links used for state and intent sharing between aircraft were assumed to be operating nominally, with sufficient bandwidth to accommodate all data transfers, and with perfect message reception. An informal internal study suggests bandwidth will be available to handle UML-4 traffic levels if an Efficient Universal Trajectory Language (EUTL) format is used.
- Because full system behaviors are not completely understood, the technology needed to address such behaviors is not addressed. Behaviors resulting from blunders or other rule violations by traffic, losses of communication, and reduced navigation precision are examples of unknowns that should be subjects of future research.
- Success criteria were set as starting points for requirements development. Actual requirements may be much more stringent than the current criteria.

9.3 Feasibility Conclusions

The flight test results provide positive indications that a vehicle-centric implementation of FPM can be made operational in the future. FPM has the potential to address critical challenges in managing the high-density traffic operations envisioned for UML-4 and above. Airspace resource management functions, such as arrival scheduling and aircraft sequencing, are additional challenges posed by high-density operations. FPM does not address these directly, but it may provide an additional benefit by reducing their development complexity. A vehicle-centric FPM function may effectively increase control power at the system level. If aircraft have greater capability and latitude to comply with scheduling constraints, the technical challenge of maintaining those constraints in the presence of disturbances may be less daunting.

Insights gained from the flight test were used to assess current state of knowledge regarding the six foundational research questions identified in Section 5. Results follow.

1. What CR success rate should FPM be required to achieve?

   Ultimate operational approval will likely require extremely high target levels of safety. An inner tactical layer of protection is expected, so strategic CR may be responsible for a lower level of assurance than the overall system requires. Nevertheless, a LOS is an extremely rare event in current operations. System-level design research is required to determine the allocation of responsibility between the strategic and tactical layers. As a starting point, this research defined full CR success as a 90% successful strategic deconfliction rate. That success criterion was exceeded, but a much greater success rate will probably be required for operational approval.

2. What separation standards will ensure adequate traffic flow for expected UML-4 traffic density?

   Establishing standards for minimum separation in UML-4 operations will involve extensive research using large-scale validated traffic simulations. For this research, reduced separation minima were assumed necessary to accommodate anticipated UML-4 traffic densities. Results indicate existing FPM automation and flight guidance technology are easily capable of supporting safe flight operations using these minima. It should be noted that separation standards are determined based on many factors, and this research only considered feasibility based on capabilities of FPM automation and aircraft guidance systems. Failure scenarios were not investigated. The research also assumed a 4D guidance capability would be used. A 4D guidance approach may not be adopted in the future.

3. At UML-4 traffic densities, will human operator attention and reaction time (and their variability) be adequate to allow humans to be in decision loops of the FPM function?

   Initial results suggest human operators residing on the aircraft may have enough time to be in the path replanning decision loop. Pilots were instructed to select a resolution advisory from the first computation cycle if possible (20 seconds default unless explicitly configured), but to use more time if necessary. There were several instances where CRs were executed from a later cycle. Post-flight analysis indicates the later resolutions were of acceptable quality. The default CD horizon of three minutes may therefore be larger than necessary. Nevertheless, longer CD horizons may allow more time for decision-making. Maneuvers involving increased time parameter conditions were observed to be more aligned with the natural tempo of operations conducted by the pilots. During debriefing sessions, research pilots directly stated a desire for longer resolution cycles, mentioning that most resolution sets required more time to evaluate adequately. It should be noted that in the test, pilots were intentionally given no recommendations by the automation to aid them in their selections. The research pilots were also focused on FPM as their primary function during test runs. Results may have been different if the FPM task had been added to their typical task load. Research to date has not investigated the ability of remote human operators to be in the decision loop.

4. What is the minimum traffic separation (separation standard plus any necessary buffers) required by FPM to achieve acceptable CD and CR performance?

   The answer to this question is critical in assessing realism of the long-term Urban Air Mobility (UAM) vision. Large separation minima will limit system capacity. Research to date has provided only an initial exploration of sizing for separation minima and buffers. Trajectory Prediction Uncertainty Bounds (TPUBs) were sized based on the trajectory conformance capabilities of the two test aircraft. Trajectory conformance precision for both aircraft was found to exceed expectations significantly. Low trajectory conformance was anticipated and was a factor in selecting the small separation minima used in the research. This foundational technical question can only be answered fully with an understanding of trajectory conformance precision required by the concept of operations.

5. Can FPM technology (automation plus associated air-to-air communications) adequately avoid triggering a domino effect?

   Chain reaction conflicts representative of the domino effect phenomenon have not manifested in any FPM research activities to date. Flight test results also indicate that the AOP CP function is effective at UML-4 traffic densities. Comprehensive testing in simulations, with each aircraft in the simulation having FPM capability, will be necessary to verify this initial conclusion. Flight test stressor maneuvers that increased traffic density were designed to reveal domino effect behaviors (such as secondary conflicts beyond the CR horizon), but none were found in those maneuvers. The traffic generator's limitations in creating extremely high traffic densities narrowed the scope of the stressor scenario. The CP function could create adverse effects. Because AOP rejects any solution that would create a conflict, AOP may provide no solutions in some situations. This was not observed. Several approaches have been identified for reducing such adverse effects, including increasing the number of AOP's resolution patterns, modifying CP time horizons, and incorporating a relaxation strategy. Initial results indicate none of these are necessary for UML-4 traffic densities and operations using the assumed separation minima.

6. How do FPM functions designed for unconstrained airspace perform in constrained airspace such as UAM corridors?

   Research to date indicates dense traffic in highly constrained airspace, such as flow corridors, will require specialized FPM functionality. Observations from flight test data show that FPM functions are capable of handling corridor operations in a manner similar to how operations are handled in open airspace, but with a reduced number of resolution options. However, design and simulation of the test maneuvers in preparation for the flight test revealed current AOP functionality was not adequate for many traffic scenarios. Although some observed behaviors were caused by known AOP defects, further exploration in validated simulation will be required to gain a deeper understanding of functional needs to support corridor operations. Dynamic path planning may need to be combined with in-trail spacing functions. Of the maneuvers flown, one scenario partially demonstrated that aircraft containing an RTA at a destination within a corridor could merge at the corridor entrance and fly without conflicts to the destination. A corridor-constrained head-on scenario was also successful, while an in-trail corridor scenario was not.

10 Future Work

Based on the findings from this work, several key recommendations emerge regarding future research on FPM automation in UAM operations. The positive results from the flight test activity suggest further investment into research and development of FPM automation is necessary to support routine high-density airspace operations. Next steps should focus on conducting extensive simulation validation, refining the research technology prototype (AOP or similar) and performing functional redesign and integration of FPM automation functional capabilities into a future flight management system (FMS) architecture. Additionally, further testing in operational environments will be essential to validate performance and scalability. Collaboration with industry and regulatory bodies should also be prioritized to ensure aspects of practical implementation. The following sections provide detailed recommendations for continued research. Some recommendations support the next steps mentioned above. Others address needs for long-term research.

10.1 Key Recommendations

- Conduct further simulation validation and model enhancements to increase confidence in future research results.

- Refine the research technology prototype to support future research.

- Identify roles and responsibilities for the human operator and develop a human/machine interface. Conduct human-in-the-loop research to validate and refine human operator roles.

- Conduct research in simulation using the refined research prototype technology. Refine and expand stressor scenarios. Expand research topics to include multi-aircraft interactions. Execute large-scale batch simulations to generate statistically significant data sets. Make use of the success criteria defined in this activity. For all future simulation experiments, control and document computational performance of the platform that hosts FPM technology.

- Continue flight testing to provide data for concept validation and technology refinement. Future flight tests should require at least two aircraft be equipped with the reference prototype technology. They should include a range of flight guidance options, including those that relax the ground-referenced 4D guidance assumption. They should test an integrated human/machine teaming concept and a supporting interface. They should also test a concept for interoperability between strategic and tactical deconfliction systems. If possible, they should include the ground-based scheduling component of an integrated air/ground system. Unscripted scenarios should also be included if they can be made safe in a flight test environment. Unscripted scenarios introduce events, triggered manually and externally to the ownship, that ensure some degree of unpredictability to the technology and to researchers.

- Perform a functional and software redesign that integrates AOP functions into a future FMS architecture. The experience of adapting AOP for future UAM vehicles revealed shortcomings in previous architectural assumptions. AOP's dynamic path planning functions were originally designed to operate in concert with a commercial transport FMS. AOP designers assumed future FMS systems would continue to be responsible for management of the active and alternate ("mod") routes, have the capability to make trajectory changes to comply with RTA changes, and provide flight guidance. These functions and AOP's planning functions are interdependent. AOP's adaptation for the current activity, which did not involve reliance on external FMS functions other than flight guidance, required some FMS functions to be incorporated into AOP. The rapid incorporation revealed several challenges, as explained in the Results portion of this report. Chief among the challenges is management of the hosting aircraft's active and alternate trajectories. A redesign will correct several issues encountered in the flight test, and it will expedite the development of future airborne automation functions.

The subsequent sections elaborate on the recommendations for continued research specific to the advancement of FPM automation technology and operating environment requirements.

10.2 FPM Automation Technology Research

The following topics describe needs for future research dedicated to refinement of FPM technology, the expansion of FPM functions. Anticipated future research approaches include 1) large-scale batch simulations that use representations of many FPM technology-equipped aircraft and their supporting systems; 2) human-in-the-loop simulations supported by to-be-developed operator interfaces; and 3) theoretical research that leads to generalizable conclusions.

- Determine the effect of TPUBs sizing on false detections and missed detections. The research will require simulations to have an adjustable representation of actual navigation performance for all simulated aircraft.

- Characterize quality and usability of the ownship trajectories generated by AOP as a function of wind prediction error. Consider using the ability to comply with an RTA as a measure of usability.

- Characterize quality and usability of the ownship trajectories generated by AOP as a function of AOP's performance model accuracy and sophistication. Consider using the ability to comply with an RTA as a measure of usability.

- Continue to explore the availability of strategic CRs.
  - Determine the number of resolution advisories available as a function of traffic density.
  - Determine the number of resolution advisories available as a function of time to first LOS (TTFLOS). Analyze results and draw conclusions relative to the existing priority rules levels.
  - Determine the number of resolution advisories available as a function of the fraction of restricted airspace to total airspace.
  - Determine the number of resolution advisories available as a function of airspace corridor dimensions.
- Explore approaches to reduce or eliminate instances of conflicts AOP cannot resolve.
  - Determine the need or benefits achievable from additional AOP Pattern-Based Genetic Algorithm (PBGA) patterns, either to mitigate unresolvable conflict instances or to add capability. If necessary, determine computational trade-offs and limits associated with the addition of complex patterns.
  - Identify and explore alternative approaches to the addition of new PBGA patterns. For example, determine the effectiveness and drawbacks (such as delay) associated with combining patterns in multiple resolution executions (which would require the presence of a constraint relaxation function).
  - Determine the need and investigate alternative implementations of strategic resolution, including combining strategic approaches with other approaches such as airborne spacing (also known as interval management). Such alternative implementations may be needed for operation within flow corridors.
- Evaluate acceptability of resolution trajectories to human operators. Acceptability research may include flyability of the trajectories, investigation of need for modifications or additions to the current resolution degrees of freedom, need for additional acceptability logic or rules, and the consistency or stability of advisories across CR refresh cycles.
- Evaluate PBGA resolution optimization characteristics and refine the technology if necessary. All optimization research should consider the relationships of fitness function optimization criteria and their relative weightings to the flight's mission objectives.
  - Analyze resolution trajectories based on one or more specified optimization measures, such as path length, energy/fuel, or time delay.
  - Explore optimization algorithm refinement strategies through variation of PBGA parameters such as population size, number of generations, pattern design parameters, and optimization criteria, tolerances, and weightings.
  - Determine the fraction of resolution advisories that, based on the optimization measures, improve the original active trajectory in UML-4 operations. If the fraction is significant, further benefits may be achieved by executing PBGA periodically, in addition to executing from an event trigger such as a conflict.

10.3 UML-4 Requirements Research

The following topics describe needs for future research to facilitate development of functional requirements for an FPM system that can support UML-4 operations. Such requirements research will involve the understanding of operational behaviors at the airspace system level. The research approaches listed for FPM Automation Technology Research will likely be needed.

- Determine system-level behaviors that result from relaxing the test design requirement for 4D vehicle guidance. Evaluate system effectiveness, system efficiency, system robustness, and operator workloads resulting from use of relaxed guidance approaches such as open-loop guidance for the time dimension, a periodic update of intent based on exceeding specified path tolerances, and other methods. Based on results, establish actual navigation performance needs for future UAM vehicles operating in UML-4 environments. Also identify system characteristics that lead to requirements for supporting systems, such as the impacts of frequency of shared intent updates on communication requirements.
- Determine system-level behaviors resulting from reduced levels of shared intent. Evaluate system effectiveness, system efficiency, system robustness, and operator workloads resulting from use of limited intent content based on time horizons, simplified trajectory representations such as use of a small number of simplified trajectory change points, and other methods. Determine minimum intent exchange requirements for UML-4 operations.
- Establish automation technology time parameters that optimize FPM system performance within a UML-4 traffic density environment. Time parameters include CD and CR lookahead horizons, CR freeze horizons, and resolution computation refresh cycle times.
  - Determine impacts of time parameter settings on flight objectives, to include optimization metrics (energy/fuel, time, ride quality, etc.) and effectiveness metrics (missed/false alerts, resolutions availability, number of resolutions required, etc.).
  - Characterize the ability of strategic resolution to meet an RTA as a function of RTA change magnitude and the time to RTA.
- Evaluate system stability at UML-4 and greater traffic densities. Include the exploration of trade-offs such as CP levels (adjustable through CR look-ahead settings), domino effect, and their interdependencies.
  - Determine the number of conflicts that occur after executed resolutions as a function of CR horizon. This is a measure of the utility of CP at the airspace system level.
  - Develop an airspace entropy parameter that quantifies the performance of the CP function and provides an indication of overall system stability. It will likely be based on the number of new conflicts that occur as a result of a CR. A candidate non-dimensional measure may be the number of conflicts that occur after a CR divided by the number of total conflicts. For simulation evaluations, the domino effect parameter of Reference 27 may be appropriate.
  - Develop a CR stability parameter that quantifies the stability of the strategic CR function. It will indicate how frequently an executed resolution advisory fails to resolve a conflict with a specific aircraft, thereby causing multiple conflicts and multiple resolutions with the same traffic aircraft.
- Establish system-level effectiveness of priority (right-of-way) rules and their associated alerting levels and alerting time parameters for a UML-4 traffic density environment.

- Establish system-level time characteristics to support development of alerting and advisory timing requirements for human decision making within the UML-4 environment. System time characteristics knowledge will facilitate future investigations of the feasibility of human-in-the-loop or human-on-the-loop decision-making, or it may establish a requirement for decision making to be an automated function.

- Evaluate FPM automation performance when also constrained by potential coordination procedures with external functions such airspace management services. Determine performance degradation as a function of time delays between automated CR recommendations and executions of the recommendations. The time delays can originate from multiple sources, including delays associated with service provider decision-making, delays in transmission of data to and from a source external to the aircraft, and losses of communication links.

- Establish vehicle-to-vehicle and vehicle-to-ground communication bandwidth requirements for a UML-4 traffic density environment.

- Consider and explore the potential to include automated executive functions that make path planning or mission planning decisions based on an assessment of the current situation. A negotiation capability may allow aircraft to iterate with ground-based resource management functions to achieve efficient flow management solutions. Executive automation functions may also provide value from the integration of in-flight planning with other future automation functions. Integration with future vehicle health assessment functions will enable replanning due to degraded aircraft performance. Integration with functions that monitor human operator state may enhance flight safety. Integration of path and mission planning functions with fleet management functions may further enhance the mission objectives of the operator.

References

1. National Academies of Sciences, Engineering, and Medicine: Advancing Aerial Mobility: A National Blueprint. 2020.

2. Federal Aviation Administration: Urban Air Mobility (UAM) Concept of Operations, v2.0. 2023.

3. Goodrich, K. H.; and Theodore, C. R.: Description of the NASA Urban Air Mobility Maturity Level (UML) Scale. AIAA-2021-1627, 2021.

4. Sharma, V.; and Wing, D. J.: Dynamic Path Planning Automation Concept for Advanced Air Mobility. NASA TM–20220009974, 2022.

5. RTCA: RTCA Unveils Member Report on Digital Flight. 2022. URL `https://www.rtca.org/news/digital-flight-report/`.

6. Wing, D. J.; Chancey, E. T.; Politowicz, M.; and Ballin, M. G.: Achieving Resilient In-Flight Performance for Advanced Air Mobility through Simplified Vehicle Operations. AIAA-2020-2915, 2020.

7. Karr, D. A.; Wing, D. J.; Barney, T. L.; Sharma, V.; Etherington, T. J.; and Sturdy, J. L.: Initial Design Guidelines for Onboard Automation of Flight Path Management. AIAA-2021-2326, 2021.

8. Ballin, M. G.; Sharma, V.; Vivona, R. A.; Johnson, E. J.; and Ramiscal, E.: A Flight Deck Decision Support Tool for Autonomous Airborne Operations. AIAA-2002-4554, 2002.

9. Wing, D. J.; and Cotton, W. B.: Autonomous Flight Rules: A Concept for Self-Separation in U. S. Domestic Airspace. NASA TP–2011-217174, 2011.

10. Karr, D. A.; Sturdy, J. L.; Barrows, B. A.; and Ballin, M. G.: An Experimental System for Strategic Flight Path Management in Advanced Air Mobility. AIAA-2024-0710, 2024.

11. Karr, D. A.; and Vivona, R. A.: Conflict Detection Using Variable Four-Dimensional Uncertainty Bounds to Control Missed Alerts. AIAA-2006-6057, 2006.

12. Silva, C.; Johnson, W.; Antcliff, K. R.; and Patterson, M. D.: VTOL Urban Air Mobility Concept Vehicles for Technology Development. AIAA-2018-3847, Aviation Technology, Integration, and Operations Conference, 2018.

13. RTCA: DO-365 - Minimum Operational Performance Standards (MOPS) for Detect and Avoid (DAA) Systems. 2017. URL `https://my.rtca.org/productdetails?id=a1B36000003FXH3EAO`.

14. Hagen, G. E.; Guerreiro, N. M.; Maddalon, J. M.; and Butler, R. W.: An Efficient Universal Trajectory Language. NASA TM–2017-219669, 2017.

15. NASA: Expanded Guidance for NASA Systems Engineering, Volume 1: Systems Engineering Practices. 2016.

16. Barney, T. L.; and Barrows, B. A.: Initial Performance Evaluation of Flight Path Management Onboard Automation. 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC), 2022.

17. Barrows, B. A.; Ballin, M. G.; Barney, T. L.; Nelson, S. L.; Underwood, M. C.; and Wing, D. J.: Sim to Flight: Evaluating Flight Path Management Automation in High Density Urban Environments. AIAA-2024-1077, 2024.

18. Nelson, S. L.; Ballin, M. G.; Barrows, B. A.; Underwood, M. C.; Wing, D. J.; Sturdy, J. L.; and Williams, E. R.: Designing a Flight Test of a Flight Path Management System for Advanced Air Mobility Research. AIAA-2024-1076, 2024.

19. Rimjha, M.; Li, M.; Hinze, S.; Tarafdar, S.; Hotle, S.; Swingle, H.; and Trani, A.: Demand Forecast Model Development and Scenarios Generation for Urban Air Mobility Operations. NASA NIA Contract NNL13AA08B, 2021.

20. NASA: AFPD-8700.1-001E Armstrong Flight Policy Directive. Available: `https://www.nasa.gov/wp-content/uploads/2021/08/afpd-8700.1-001_0.pdf`, 2021.

21. Baccheschi, N.; Eng, D.; Guion, A.; Nwajagu, E.; Ramia, S.; Scofield, J.; Williams, E.; Fettrow, T.; and Monk, K.: Integration of Automated Systems Test Campaign. NASA TM-20240002679, 2024.

22. International, A.: ASTM F3269-21 Standard Practice for Methods to Safely Bound Behavior of Aircraft Systems Containing Complex Functions Using Run-Time Assurance. 2021.

23. Karr, D. A.; Vivona, R. A.; Woods, S. E.; and Wing, D. J.: Point-Mass Aircraft Trajectory Prediction Using a Hierarchical, Highly-Adaptable Software Design. AIAA-2017-3336, 2017.

24. Wing, D. J.: The TASAR Project: Launching Aviation on an Optimized Route Toward Aircraft Autonomy. NASA TP–2019-220432, 2019.

25. Karr, D. A.; Vivona, R. A.; Roscoe, D. A.; DePascale, S. M.; and Consiglio, M.: Experimental Performance of a Genetic Algorithm for Airborne Strategic Conflict Resolution. AIAA-2009-5748, 2009.

26. Monk, K. J.; Fettrow, T. D.; Galvan-Garza, R. C.; and Kraft, A. E.: Physiological and Subjective Responses of Pilots during Advanced Air Mobility Flight Testing with Automated Systems. NASA/TP–20240007669, 2024.

27. Krozel, J.; Peters, M.; Bilimoria, K. D.; Lee, C.; and Mitchell, J. S. B.: System Performance Characteristics of Centralized and Decentralized Air Traffic Separation Strategies. 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, NM, 2001.

Appendix A: Performance Measures and Metrics

The following descriptions of the measures of performance (MOPs) and metrics were used in the analysis of flight test results.

A.1 Measures of Performance

MOP 1: There is a low fraction of false conflict alerts (false positives) and a very low fraction of missed conflict alerts (false negatives).

- Success Criteria:
  i. 5 percent or fewer missed detections. The ratio (number of missed detections)/ (number of true detections + number of missed detections) is less than 0.05.
  ii. 20 percent or fewer false detections. The ratio (number of false detections)/ (number of true detections + number of false detections) is less than 0.20.
- Rationale for Success Criteria: Engineering judgment based on concept assumptions. There will likely be inner layers of CD with shorter time horizons, provided by tactical collision avoidance systems and/or future FPM functions. The total fraction of conflicts detected and resolved by the outer layer may be required to be large, possibly greater than 99%. As a starting point for further research, the missed conflict alerts criterion for the current study sets a requirement for FPM to allow no more than 5% of all conflicts to be detected by inner tactical layers. TPUBs are used to ensure missed conflict alerts are low, but their use may result in false conflict alerts. False conflict alerts may unnecessarily impact traffic flow, flight efficiency, or operator workload. The false conflict alerts criterion of 20% is a starting point for further research. The criteria may serve as a basis for a future design requirement.
- Analysis Approach: All encounters that did not involve the execution of a resolution will be analyzed. Time and separation distance at the closest point of approach (CPA) will be predicted by the technology at the first CD. These values will be compared with the actual time and separation distance at the CPA. For encounters involving two real aircraft, estimates of the actual values will be required for test maneuvers that end prior to CPA. Each estimate will be based on a prediction of CPA based on actual aircraft trajectories up to the end of the maneuver. For maneuvers that achieve CPA with another aircraft (which may happen with virtual background aircraft), actual measures will be used. Actual separation losses with background aircraft that were not detected will be considered missed detections. Each encounter not executing a resolution will be categorized as a correct detection, a missed detection, or a false detection. Analysis will be performed for lateral, vertical, and along-path CPAs, as applicable for the encounter. Analysis will be performed for the combined set of all conflicts that occur in testing as well as individually for level flight conflicts, climbing conflicts, descending conflicts.

  Test data alone will probably be too sparse to evaluate the success criteria statistically. In addition, test maneuvers are designed to create conflicts. If maneuvers are designed successfully, they may prevent the measurement of a false detection rate. False detections for the ownship would only be possible with virtual background traffic. MOP 1 criteria may prove to be relevant only in future research that uses validated traffic simulations, which are beyond the scope of the current test activity.

- Relevant Measures of effectiveness (MOEs): 1a, 1b, 1c, 3, 4

MOP 2: Conflicts are detected within the first two CD cycles, measured from when first detection should have occurred.

- Success Criterion: True for at least 90 percent of all occurrences
- Rationale for Success Criterion: Engineering judgment based on concept assumptions. Conflicts must be detected early enough to allow for the CR function, operators, and other system agents to respond. Each detection cycle within AOP is of three seconds duration. Two detection cycles allow six seconds for detection. The 90% value is an initial baseline, to be used as a starting point for further research to determine an appropriate target. The criterion may serve as a basis for a future design requirement.
- Analysis Approach: For each encounter, a time when the first conflict alert should have occurred is determined. This time is either the CD lookahead horizon or the time that a triggering event occurs, such as an intent change by another aircraft inside the lookahead horizon. The time is compared with the actual time of the first conflict alert. Analysis will be performed for conflicts with other aircraft and with special use airspace (SUA) boundaries, if the latter is possible from collected data. Analysis will be performed for the combined set of all conflicts that occur in testing as well as individually for level flight conflicts, climbing conflicts, descending conflicts. Test data alone will probably be too sparse to evaluate the success criterion statistically. Analysis may require use of validated simulation, which is beyond the scope of the current test activity.
- Relevant MOEs: 1a, 1b, 2a, 3

MOP 3: At least one CR advisory is provided for each detected conflict within the first CR refresh cycle.

- Success Criterion: True for 90 percent of all occurrences
- Rationale for Success Criterion: Engineering judgment based on the FPM concept of operations. There is no guarantee of a solution, so the success criterion states that at least one solution must be found. Further, for priority rules to function, resolution trajectories must be found early enough to allow for CR execution in the earliest alerting period. The number of solutions available is expected to reduce as the TTFLOS reduces. The CR refresh cycle time is used as a measure of time that, depending on configuration settings, is likely to vary proportionally with other varied horizon configuration settings. The value of 90 percent is set as a starting point for further research. The criterion may serve as a basis for a future design requirement.
- Analysis Approach: All encounters between the ownship and traffic and between the ownship and restricted airspace are evaluated, whether intended or not, in the maneuver design. For each set of resolution advisories, determine the time between the initial conflict alert and when at least one resolution advisory is provided. Log cases for which no resolution is provided. Determine what fraction of the total occurs within the specified time. Test data alone will probably be too sparse to evaluate the success criterion statistically. Analysis may require use of validated simulation, which is beyond the scope of the current test activity.
- Relevant MOEs: 1a, 1b, 1c, 3

MOP 4: Over the design range of CR validity, executed resolutions result in no LOS between the resolving aircraft and any conflicting aircraft.

- Success Criteria:
  i. Executed resolutions meet or exceed minimum required separation at CPA between the ownship and all traffic and restricted airspace for at least 90% of detected conflicts.
  ii. 100% of executed resolutions are flyable within the hosting aircraft's performance constraints. The design range of CR validity is the time for which CR should be producing a conflict-free trajectory advisory. It is a time window bounded by the hosting aircraft's current position and the CR horizon.
- Rationale for Success Criteria: Engineering judgment. For a mature system, a conflict that is detected by an FPM system is expected to be resolved by the system. Accounting for the lower level of maturity of the reference technology, the successful resolutions criterion is set at 90%. This is an initial baseline, to be used as a starting point for further research. All resolutions must be flyable by the aircraft, so the second criterion is set at 100%. The two criteria may serve as a basis for a future design requirement.
- Analysis Approach: Only executed resolutions will be analyzed. Each resolution encounter will be categorized as a successful resolution or a failed resolution. The CPA for the executed resolution advisory will be estimated based on a prediction of the remaining trajectory from the end of the test run. Any actual CPA measures available (such as for conflicts with virtual background aircraft) will be used. Flyability of the resolution will be determined by comparing the flown trajectory with trajectory constraints AOP uses in trajectory generation, such as turn rate, climb rate, and speed limits. Analysis will be performed for lateral, vertical, and along-path CPAs, as applicable for the encounter. Analysis will be performed for the combined set of all executed resolutions that occur in testing as well as individually for level flight conflicts, climbing conflicts, descending conflicts. Test data alone will probably be too sparse to evaluate the success criteria statistically. Analysis may require use of validated simulation, which is beyond the scope of the current test activity. If time allows, test data will be used for simulation validation.
- Relevant MOEs: 1a, 1b, 1c, 3

MOP 5: CRs do not result in a repeat conflict between the resolving aircraft and the original conflict aircraft within the design range of CR validity.

- Success Criterion: Repeat conflict alerts do not occur for at least 90% of resolved conflicts. Repeat conflict alerts are defined as conflicts detected with one or more aircraft for which a conflict has previously been detected and a CR has been executed. The design range of CR validity is the time for which CR should be producing a conflict-free trajectory advisory. It is a time window bounded by the hosting aircraft's current position and the CR horizon.
- Rationale for Success Criterion: Engineering judgment based on the FPM concept of operations. Once a CR has been executed, the detected conflict should not reappear within the CR look-ahead horizon if the traffic aircraft in conflict conforms with its shared trajectory intent and does not change its intent. This system characteristic may prove critical to acceptable system stability and crew workload. The success criterion is an initial baseline, to be used as a starting point for further research. The criterion may serve as a basis for a future design requirement.
- Analysis Approach: Only executed resolutions will be analyzed. Any resolution that results in a repeat detection within the design range of CR validity will be categorized as having failed. Analysis will be performed for the combined set of all executed resolutions that occur in testing. If observed, conditions associated with failures will be documented. Test data alone will probably

be too sparse to evaluate the success criterion statistically. Analysis may require use of validated simulation, which is beyond the scope of the current test activity. If time allows, test data will be used for simulation validation.

– Relevant MOEs: 1a, 2a, 2b, 3

MOP 6: CRs will not create new traffic or airspace conflicts within the design range of CR validity.

– Success Criterion: At least 90% of executed CRs create no new conflicts with traffic or airspace regions over the design range of CR validity, excluding conflicts caused by a change of traffic intent within the CR horizon.

  The design range of CR validity is the time for which CR should be producing a conflict-free trajectory advisory. It is a time window bounded by the hosting aircraft's current position and the CR horizon.

– Rationale for Success Criterion: Engineering judgment. System stability may depend on CP to reduce or eliminate a domino effect. The CP success criterion is an initial baseline, to be used as a starting point for further research. The criterion may serve as a basis for a future design requirement.

– Analysis Approach: Only executed resolutions will be analyzed. Any resolution that results in a new conflict with any traffic aircraft or SUA region within the design range of resolution validity will be categorized as a failed instance of CP. Analysis will be performed for the combined set of all executed resolutions that occur in testing. If observed, conditions associated with failures will be documented.

  Test data alone will probably be too sparse to evaluate the success criterion statistically. Analysis may require use of validated simulation, which is beyond the scope of the current test activity. If time allows, test data will be used for simulation validation.

– Relevant MOEs: 1a, 1b, 1c, 2b, 3

MOP 7: CRs will meet an RTA constraint within aircraft performance limits.

– Success Criterion: At least 90% of all computed CRs from the first CR cycle will meet an existing RTA within three seconds if the point of RTA is at least three minutes downstream the hosting aircraft's current position.

– Rationale for Success Criteria: The future UML-4 concept may rely on some form of metering or time-based flow management. If so, the ability of the aircraft to comply with an assigned arrival time while avoiding traffic and complying with airspace constraints will be critical. The success criterion is based on engineering judgment and AOP design settings. The three seconds precision is based on the current AOP tolerance setting. The point of RTA (3 minutes downstream) is based on engineering judgment in the absence of simulation results. The success percentage number is an initial baseline, to be used as a starting point for further research. The criterion may serve as a basis for a future design requirement.

– Analysis Approach: RTA trajectory conformance will first be validated using a small number of runs that execute a resolution and allow the ownship to fly close to the point of RTA. Once trajectory conformance is validated, both executed resolutions and advisories not executed will be analyzed as a combined set using the predicted RTA crossing times from the resolution advisories. RTA crossing times will be compared with the RTA values. Test data alone will probably be too sparse to evaluate the success criterion statistically. Analysis may require use of validated simulation, which is beyond the scope of the current test activity. If time allows, test data will be used for simulation validation.

– Relevant MOEs: 1c, 3

MOP 8: FPM technology will prioritize CRs over RTA constraints. If necessary to resolve a traffic conflict, it will provide a CR solution that is out of compliance with the RTA, but as close to compliance as possible.

– Success Criteria: If an RTA cannot be met while resolving a traffic conflict.
  i. The conflict will be successfully resolved for 90 percent of all occurrences.
  ii. The RTA constraint will be relaxed by the minimum amount necessary for the hosting aircraft to remain within performance constraints for 90 percent of all occurrences.

– Rationale for Success Criteria: CR has priority over RTA compliance. Rather than providing no solutions, AOP CR will relax the RTA constraint to increase the probability that a solution is available to resolve traffic conflicts. To meet the RTA as closely as possible, AOP will either provide a solution that results in the aircraft operating at its performance limits, and/or it will provide a solution that provides minimum separation from the conflicting aircraft. A value of 90 percent was chosen as a starting point because it is applied to other CR success criteria. The criterion may serve as a basis for a future design requirement.

– Analysis Approach: Test runs involving conflicts that result in RTA noncompliance will be analyzed to determine if RTA constraint relaxation was necessary. Constraint relaxation is indicated if the resolution trajectory involves operating at the hosting aircraft's performance limits, and/or depending on situation, the resolution provides minimum separation from the conflicting aircraft. These resolution trajectory characteristics will be considered evidence that a minimum level of constraint relaxation has been applied by PBGA. Test data alone will be too sparse to evaluate the success criteria statistically, so analysis will only consist of results observations. Statistical analysis will require use of validated simulation, which is beyond the scope of the current test activity. If time allows, test data will be used for simulation validation.

– Relevant MOEs: 1c, 3

MOP 9: FPM technology will resolve traffic conflicts in airspace corridors of width between 0.5 and 1.0 NM that are not constrained by RTAs located within the corridor, at expected UML-4 density.

– Success Criteria:
  i. Conflicts involving corridor operations are resolvable for at least 90% of detected conflicts.
  ii. Executed resolutions meet or exceed minimum required separation at CPA between the ownship and all traffic and restricted airspace for at least 90% of detected conflicts.

– Rationale for Success Criteria: The concept of FPM operations requires a very low fraction of failed CRs in airspace corridors. The current test only considered cases with RTAs located outside of a corridor. The reference automation is not designed for cases where an RTA is within a corridor. Those cases are not anticipated to be successful. The successful resolutions criterion is an initial baseline, and is set to be equal to the criterion for resolutions in open airspace. The criteria may serve as a basis for a future design requirement.

- Analysis Approach: Only executed resolutions will be analyzed. Each resolution encounter will be categorized as a successful resolution or a failed resolution. The CPA for the executed resolution advisory will be estimated based on a prediction of the remaining trajectory from the end of the test run. Any actual CPA measures available (such as for conflicts with virtual background aircraft) will be used. Test data alone will be too sparse to evaluate the success criteria statistically, so analysis will only consist of results observations. Statistical analysis will require use of validated simulation, which is beyond the scope of the current test activity. If time allows, test data will be used for simulation validation.
- Relevant MOEs: 3, 4

MOP 10: FPM technology will complete all CR computations and display CR alerts within the first 15 seconds of each refresh cycle at UML-4 traffic density.

- Success Criterion: Complete CR computations and display resolution advisories within 15 seconds for 100 percent of all occurrences.
- Rationale for Success Criterion: PBGA CR computations must complete within a single AOP CR refresh cycle. The baseline refresh cycle setting is 20 seconds. The success criterion was set to accommodate the 15-second CR refresh cycle of the minimum time parameter case to be tested. The criterion is an initial baseline only, to be used as a starting point for further research to determine an appropriate target. The criterion may serve as a basis for a future design requirement.
- Analysis Approach: For each CR cycle, the time when the cycle is triggered is subtracted from the time CRs are displayed to the operator. Results will be analyzed statistically for the entire test data set. Any outliers will be studied to gain insights regarding causes of slow computations.
- Relevant MOEs: 2a, 3

MOP 11: FPM technology will provide operators with sufficient time to evaluate and execute a resolution before a new set of resolution advisories become available to evaluate.

- Success Criterion: For maneuvers using baseline time parameters, resolution advisories are displayed to the operator for execution for at least 15 seconds before being removed or replaced, for 90% of all cases.
- Rationale for Success Criterion: Computational power must be great enough to 1) complete CR advisories computations, 2) provide advisories to a human operator, and 3) allow time for the operator to evaluate advisories, select one, and execute. The display duration success criterion was set based on engineering judgment in the absence of prior human factors experimental results. Display duration is affected by the time required for resolution computation, which can vary. On average, the time available for the operator to evaluate and execute a resolution should equal to the CR refresh cycle. The time available is the CR cycle time minus the current cycle's CR computation time plus the following cycle's CR computation time. The minimum value of 15 seconds was set based on a reasonable minimum that can reliably be achieved for the baseline time parameters test case. The 90% value is a starting point for future research. Future studies producing a larger amount of data may support a higher percentage value. The criterion may serve as a basis for a future design requirement.

- Analysis Approach: All maneuvers using baseline time parameters are evaluated. For each CR cycle, the time when CR advisories are displayed to the operator is subtracted from the time when advisories are removed or replaced with new advisories. Results will be analyzed to obtain average values over selected test sets.
- Relevant MOEs: 2a, 3

A.2 Metrics

- Metric 1: False Detection Rate

  For all conflicts without executed resolutions, compare first detection conflict time and predicted CPA separation distance with actual time and CPAs to determine if a conflict actually existed.

- Metric 2: Missed Detection Rate

  For all encounters between the ownship and traffic, assess actual CPA distance to determine if a conflict actually existed.

- Metric 3: Actual vs. Predicted CPA Error

  For all executed resolutions, compute the average deviations between the predicted CPA of resolution trajectories and actual achieved CPA.

- Metric 4: Conflict Detection Alert Delay

  For all conflicts, compute the time difference between the initial conflict alert and the triggering event. A triggering event may occur at the CD look ahead horizon or be caused by an external trigger.

- Metric 5: Rate of Resolvable Conflicts

  For all conflicts, determine the rate of occurrence for which at least one resolution was provided before end of run, typically 30 seconds prior to LOS or a "knock-it-off" call.

- Metric 6: Same-Traffic Conflict Occurrence Rate

  For all conflicts with executed resolutions, determine if a second conflict occurred with the Intruder traffic from the previous conflict.

- Metric 7: Actual LOS Rate after CR

  For all conflicts with executed resolutions, determine the percentage of LOS that occurred when flown (whether AOP detected it or not) via observed state.

- Metric 8: Rate of First CR Cycle Resolution for All Conflicts

  For all conflicts, calculate the rate of occurrence where at least one resolution was provided within the first CR cycle.

- Metric 9: Rate of First CR Cycle Resolution for Resolvable Conflicts

  For all conflicts where at least one resolution is found, calculate the rate of occurrence where at least one resolution is found within the first CR cycle.

- Metric 10: RTA Conformance Rate for Traffic Conflict Resolutions

  For all executed CRs (triggered by traffic conflicts), determine the fraction of RTA conformance magnitudes that are within the maximum acceptable time tolerance.

- Metric 11: Conflict Prevention Rate

  For all conflicts with executed resolutions, compute the occurrence rate of future conflicts with any traffic caused by the resolution of an existing conflict. Future conflicts outside the CR horizon at the time of CR computation are not considered by the metric.

- Metric 12: RTA Conformance with RTA Change Requests and Traffic Conflicts

  For all conflicts with executed resolutions, compute the average magnitude of RTA deviation caused by an RTA change request and a traffic CR.

- Metric 13: Actual vs. Predicted CPA Error in Corridor Operations

  For all conflicts involving corridor operations with executed resolutions, compute the difference between actual vs. predicted CPA after CR.

- Metric 14: Conflict Prevention Rate in Corridor Operations

  For all conflicts involving corridor operations, a) compute the occurrence rate of future conflicts with any traffic caused by the resolution of an existing conflict, and b) determine the rate of resolvable conflicts involving corridor operations. Future conflicts outside the CR horizon at the time of CR computation are not considered by the metric.

- Metric 15: CR Performance Evaluation

  Computed the elapsed time of all PBGA resolution advisories and determine the average elapsed time for the entire data set and for each resolution degree of freedom.

- Metric 16: CR Trajectory Flyability Evaluation

  For executed CR trajectories, determine whether any AOP ownship performance model limit was exceeded by any portion of the trajectory. For the flight test, performance model constraints consist of set of speed constraints and climb or descent rates.

- Metric 17: CR Trajectory Performance Limits Evaluation

  For executed CR trajectories, determine whether any portion the CR trajectory was at a limit specified by the AOP ownship performance model. For the flight test, performance model constraints consist of set of speed constraints and climb or descent rates.

- Metric 18: CR Cycle Elapsed Resolution Display Time

  For all conflicts, compute the average elapsed time between consecutive completed PBGA CR cycles. The PBGA completion event time is an estimate of the time pilot operators would first see a cycle's resolutions presented to an external display interface. The difference between two PBGA completion event times is an estimate of the time the operator has to evaluate, select, and execute resolutions.

Appendix B: FPM Maneuver Design and Constraints

All maneuvers used in the flight test include the ownship aircraft that hosts FPM technology and a simulated UML-4 airspace and traffic environment. All maneuvers involving two-aircraft separation-loss encounters include the Intruder aircraft as well. The simulated airspace and traffic environment contains recorded virtual background traffic at UML-4 density levels, restricted airspace SUA, simulated origin and destination vertiports for all aircraft, and an RTA constraint for all aircraft. Corridors are created with the use of SUA regions.

B.1 Maneuver Design Parameters

The following items are common to all test groups:

- Maneuvers were designed to comply with the following constraints:
  - Sikorsky Autonomy Research Aircraft (SARA) ownship maximum speeds: 120 kt calibrated airspeed (CAS) and 145 kt ground speed
  - SARA ownship minimum speed: 50 kt CAS
  - Optionally Piloted Vehicle (OPV) Intruder maximum speeds: 120 kt CAS and 145 kt ground speed
  - OPV intruder minimum speed: 50 kt CAS
  - Data link range limits
    * 20 NM or less distance between SARA and OPV
    * 25 NM distance from KBDR
  - Sikorsky identified an operating area over Long Island Sound where data collection would take place.
    * The rectangular area was centered immediately south of Bridgeport (KBDR), with a length of approximately 27.5 statute miles running southwest to northeast along the sound, and a width of approximately 13 statute miles.
    * The area was defined to keep aircraft within VHF coverage for each aircraft's telemetry link.
    * Pilots coordinated with KBDR tower as necessary to fly into its Class D airspace, which partially overlapped the rectangle. All test maneuvers were designed with initial trajectories outside Class D airspace. Maneuvers were also designed to have a small likelihood of resolution trajectories entering Class D airspace.
- Mission rules defined by the Integration of Automated Systems (IAS) team were followed. Detailed mission rules were contained in Document AAM-NC-098-003, "Mandatory Mission Requirements, Spiral 2C." Some highlights that impacted maneuver design include:
  - For safety, all FPM encounter geometries were designed such that the CPA of the two flight plans was never less than 0.2 NM laterally. There was no equivalent vertical offset.
  - Encounters were designed such that separation was not lost. Maneuvers were terminated well prior to separation loss. Knock-it-off (KIO) procedures were to be executed for safety when the ownship reaches a point 30 seconds prior to LOS, defined as separation less than 1500 feet laterally or 450 feet vertically.

  - If any non-participating aircraft were observed within 2000 ft vertically and 3 NM laterally of either test aircraft and could be determined to be a non-factor, then the test run was to be terminated.

- The default altitudes for both real aircraft at the start of each maneuver were be 2000 ft geometric (GPS) altitude.
- The default speeds for both real aircraft at the start of each maneuver were 93 kt CAS.
- Both real aircraft were considered equivalent to the NASA quadrotor model. EUTL trajectories for both aircraft were determined based on the performance envelope of the quadrotor model. AOP used the NASA quadrotor model to represent the ownship performance, where nominal cruise CAS for the quadrotor model is 93 kt, minimum speed is 75 kt, and maximum speed is 109 kt. Note that the model's minimum speed of 60 kt was modified to avoid operation in the speed range that caused undesired ownship oscillations due to limitations in the 4D guidance system. Lastly, nominal cruise TAS for the NASA lift-plus-cruise model is 109 kt, minimum speed is 80 kt, and maximum speed is 129 kt.
- The ownship's AOP software was configured to recognize the Intruder and all virtual background traffic as operating without AOP. As "unequipped" traffic aircraft, AOP's priority rules assigned priority to traffic. The ownship therefore yielded right-of-way to all traffic in these tests.
- By default, all virtual recorded background traffic at UML-4 density were present for all maneuvers. The background traffic were based on the Dallas Fort Worth (DFW) UML-4 traffic simulation provided by FPM.
- By default, no resolution advisories were executed after the first resolution advisory was executed. One exception is if a repeat conflict with Intruder occurred. In such a circumstance, a new CR advisory was executed.
- Maneuvers were designed to prevent ground speed limits from being encountered.
  - Data was to be treated as invalid and the test run aborted if ground speed limits were encountered.
  - AOP used an aircraft performance model to ensure it did not provide an advisory which exceeded ground speed limits of the real aircraft.
  - Maneuvers were designed such that ground speed limits would not be exceeded in the presence of zero winds or headwinds. A tailwind never-exceed limit was determined for each maneuver as part of the maneuver design. Testing was to be suspended if winds exceeded the determined limit.
  - Because prevailing wind vectors generally originated from the west, most maneuvers were designed such that the ownship traveled from east to west unless explicitly stated otherwise for a specific maneuver or circumstance.

B.2 Maneuver Descriptions

M1: Crossing encounter resulting in a traffic conflict with the intruder. The ownship and the intruder are in level cruise flight. The ownship's trajectory contains a lateral Trajectory Change Point (TCP). The intruder's trajectory contains lateral TCP. Utilized in Test Groups 1 and 2.

M2: Head-on encounter resulting in a traffic conflict with the intruder. The ownship and the intruder are in level cruise flight. The intruder's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M3: Acute crossing encounter resulting in a traffic conflict with the intruder. The ownship and the intruder are in level cruise flight. The intruder's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M4: Overtake encounter resulting in a traffic conflict with the intruder. The intruder overtakes the ownship. The ownship and the intruder are in level cruise flight. The intruder's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M5: Overtake encounter resulting in a traffic conflict with the intruder. The ownship overtakes the intruder. The ownship and the intruder are in level cruise flight. The intruder's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M6: Crossing encounter resulting in a traffic conflict with the intruder. The ownship is in level cruise flight. The intruder descends through the ownship's altitude. The ownship's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M7: Head-on encounter resulting in a traffic conflict with the intruder. The ownship is in level cruise flight. The intruder descends through the ownship's altitude. The intruder's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M8: Crossing encounter resulting in a traffic conflict with the intruder. The ownship is in level cruise flight. The intruder climbs through the ownship's altitude. The ownship's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M9: Head-on encounter resulting in a traffic conflict with the intruder. The ownship is in level cruise flight. The intruder climbs through the ownship's altitude. The intruder's trajectory contains a lateral TCP. Utilized in Test Groups 1 and 2.

M10a: RTA compliance. Straight route geometry. Large RTA delay causes AOP to produce PBGA-computed path stretches to absorb delay. Minimum speed limits not low enough to solve by speed alone. No CR execution to collect large amounts of PBGA advisory data. No intruder aircraft. Utilized in Test Group 3.

M10b: Repeat of M10a with CR execution in first cycle. Utilized in Test Group 3.

M11: [REMOVED FROM TEST PLAN] RTA compliance. Straight route geometry. Large RTA change to arrive early. Only possible to arrive on time or fail based on nature of route geometry. No CR execution to collect large amounts of PBGA advisory data. No intruder aircraft. Originally part of Test Group 3. Removed from test plan because there would be no insights gained.

M12: RTA compliance. Turn route geometry. Large RTA delay causes AOP to produce PBGA-computed path stretches to absorb delay. Minimum speed limits not low enough to solve by speed alone. No CR execution to collect large amounts of PBGA advisory data. No intruder aircraft. Utilized in Test Group 3.

M13a: RTA compliance. Turn route geometry. Large RTA change to arrive early causes AOP to produce PBGA-computed path cuts to meet RTA. Max speed limits not high enough to meet RTA by speed alone. No intruder aircraft. Utilized in Test Group 3.

M13b: Repeat of M13a with CR execution in first cycle. Utilized in Test Group 3.

M14: [REMOVED FROM TEST PLAN] RTA Compliance. Straight route geometry. RTA change to arrive earlier with "blocker" virtual traffic in front of the ownship to explore how AOP currently handles RTA changes (no PBGA). No intruder aircraft. Originally part of Test Group 3. Removed from test plan because M14 was too similar to Test Group 4 maneuvers.

M15: RTA change with intruder traffic conflict. RTA delay (approximately one minute delay) given about one minute after start and almost immediately results in a conflict with the intruder (behind ownship). CR is compliant with the new RTA. Utilized in Test Group 4.

M16: RTA change with intruder traffic conflict. RTA change to arrive early (arrive approximately one minute early) given about one minute after start and almost immediately results in a conflict with the intruder (in front of ownship). CR is compliant with the new RTA. Utilized in Test Group 4.

M17: RTA change with intruder traffic conflict. Slightly converging route geometry. RTA delay (approximately one minute delay) given about one minute after start and almost immediately results in a conflict with the intruder (behind the ownship). Conflict occurs five to six minutes from destination. The ownship is not able to meet RTA and resolve conflict (AOP must relax RTA compliance). Utilized in Test Group 4.

M18: [REMOVE FROM TEST PLAN] Arrive early case based on M17. Originally part of Test Group 4. Removed due to the difficulty involved to create the desired maneuver behavior, and due to similarity of insights gained with M17.

M19: M1 crossing maneuver with SHORT AOP time parameters. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M20: M1 crossing maneuver with LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M21: M1 crossing maneuver with VERY-LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M22: M2 head-On maneuver with SHORT AOP time parameters. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M23: M2 head-On maneuver with LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M24: M2 head-On maneuver with VERY-LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M25: M3 acute crossing maneuver with SHORT AOP time parameters. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M26: M3 acute crossing maneuver with LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M27: M3 acute crossing maneuver with VERY-LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M28: M4 intruder overtake maneuver with SHORT AOP time parameters. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M29: M4 intruder overtake maneuver with LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M30: M4 intruder overtake maneuver with VERY-LONG AOP time parameters. T_0 points may be pushed further out. Investigates effect of different AOP time parameters on CD, quality of CR advisories, number of resolutions per refresh cycle, and pilot decision making. Utilized in Test Group 5.

M31: M1 crossing maneuver with an intruder intent change (2 minutes left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M32: M1 crossing maneuver with an intruder intent change (1 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M33: M2 head-on maneuver with an intruder intent change (2 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M34: M2 head-on maneuver with an intruder intent change (1 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M35: M3 acute crossing maneuver with an intruder intent change (2 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M36: M3 acute crossing maneuver with an intruder intent change (1 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M37: M4 intruder overtake maneuver with an intruder intent change (2 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

M38: M4 intruder overtake maneuver with an intruder intent change (1 min left in CD horizon). Investigates AOP's performance when given less time until LOS. Utilized in Test Group 6.

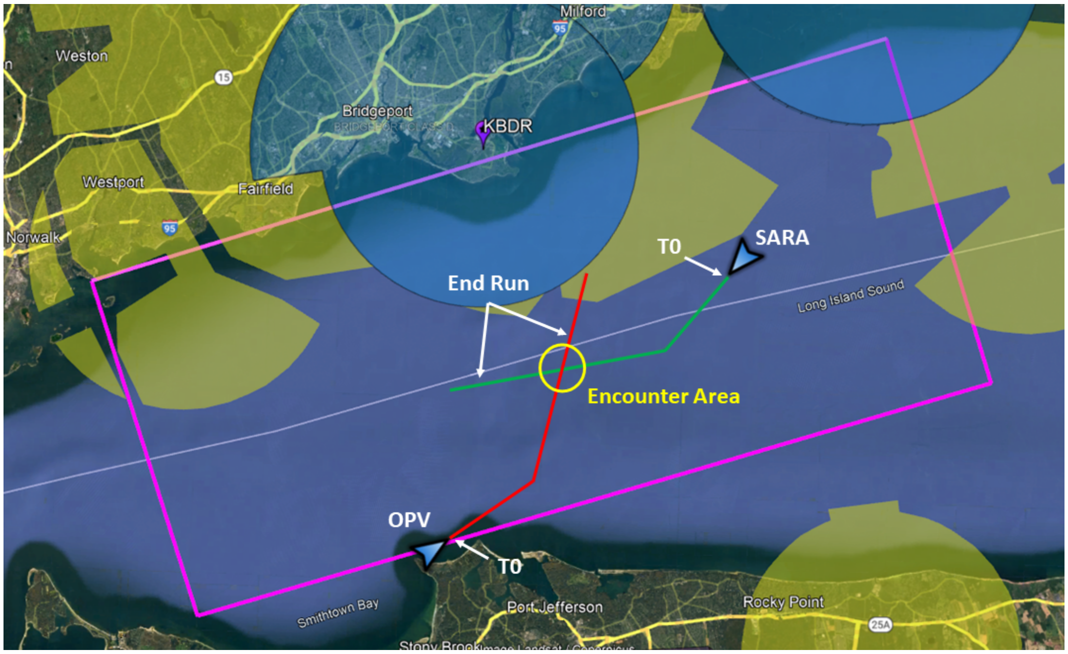


Figure B1. Maneuver 1 Crossing Geometry Over Long Island Sound Airspace (0 Winds)

M39: Corridor operations. The ownship and the intruder merge to enter DFW corridor. The ownship will land second. AOP must take action before entering the constrained corridor to avoid conflict. Utilized in Test Group 7.

M40: Corridor Operations.The intruder follows the ownship inside a corridor. The ownship receives an RTA delay that results in conflict with the intruder. The ownship is constrained by virtual background traffic above and below (altitude change not possible). The ownship must speed back up in constrained corridor and wait until exiting the corridor to handle the RTA delay. Utilized in Test Group 7.

M41: Corridor Operations. The ownship inside a corridor encounters a head-on conflict with the intruder flying inside the corridor. Likely to force the ownship to change altitude to avoid conflict. Utilized in Test Group 7.

M42: M1 crossing maneuver with approximately 25% higher traffic density (actual 1.32 times) and baseline AOP time parameters. Investigates AOP performance with traffic densities above UML-4 (as much as simulation allows). Utilized in Test Group 8.

M43: M1 crossing maneuver with 25% higher traffic density (actual 1.32 times) and LONG AOP time parameters. Investigates AOP performance with traffic densities above UML-4 (as much as simulation allows). Utilized in Test Group 8.

B.3 Core Encounter Geometries

Figures B1 through B6 represent the types of encounter geometries that were planned during the flight test.

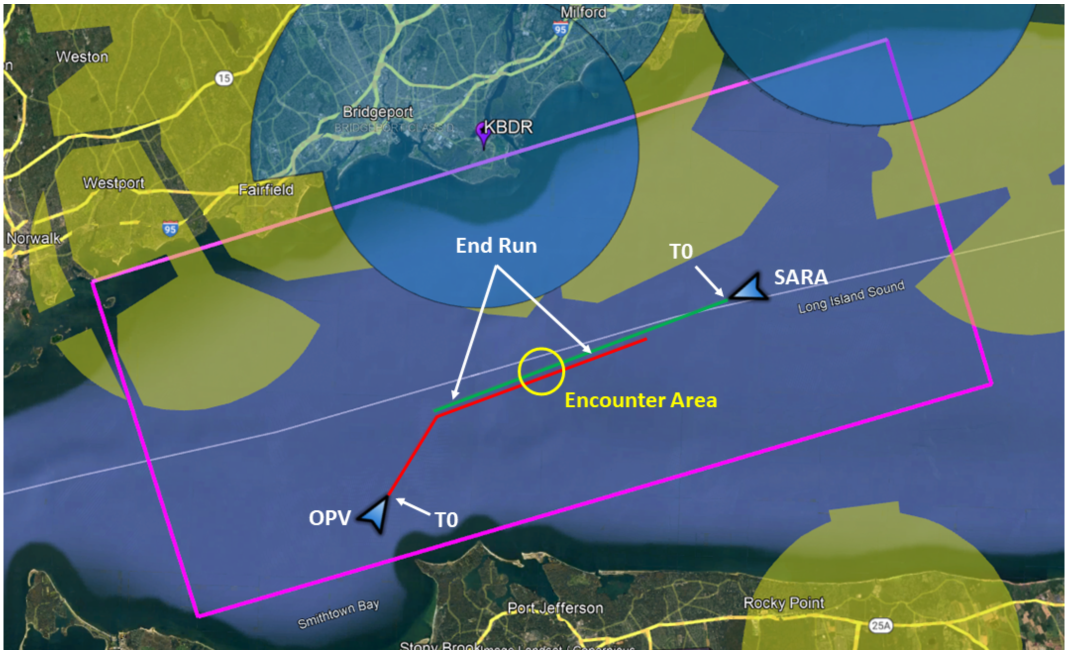


Figure B2. Maneuver 2 Head-On Geometry Over Long Island Sound Airspace (0 Winds)

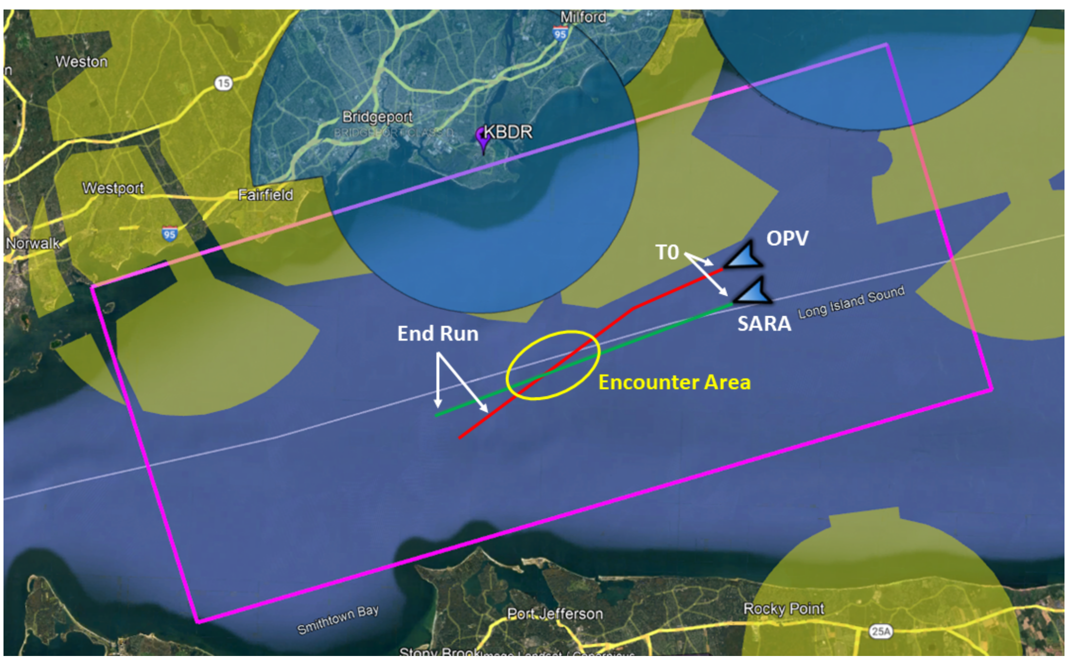


Figure B3. Maneuver 3 Acute Crossing Geometry Over Long Island Sound Airspace (0 Winds)

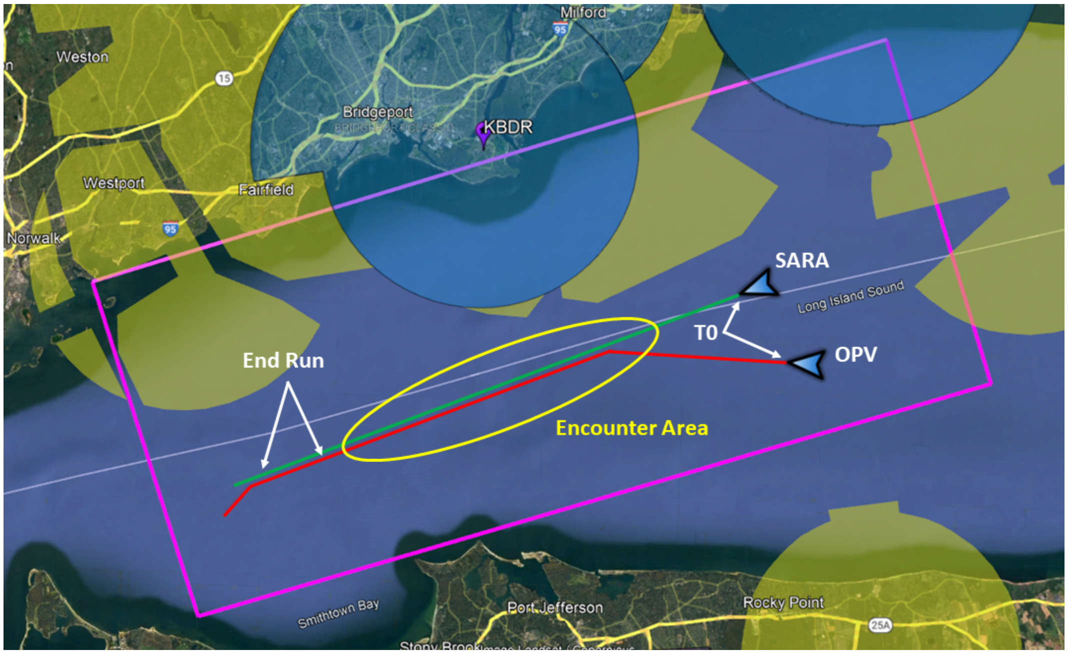


Figure B4. Maneuver 4 Intruder Overtake Geometry Over Long Island Sound Airspace (0 Winds)

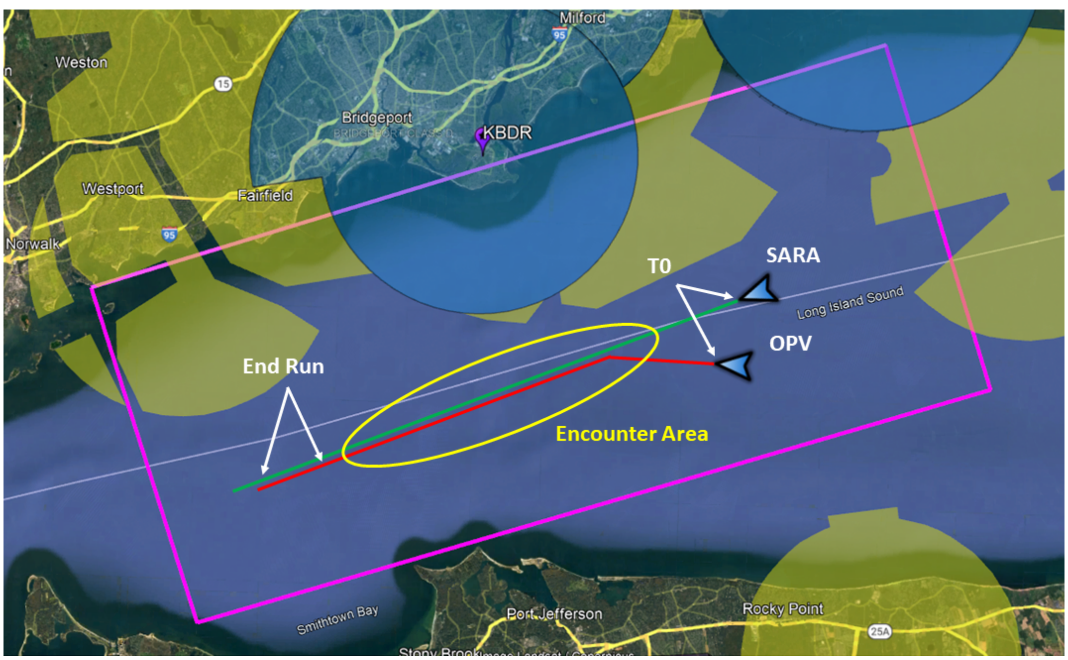


Figure B5. Maneuver 5 Ownship Overtake Geometry Over Long Island Sound Airspace (0 Winds)

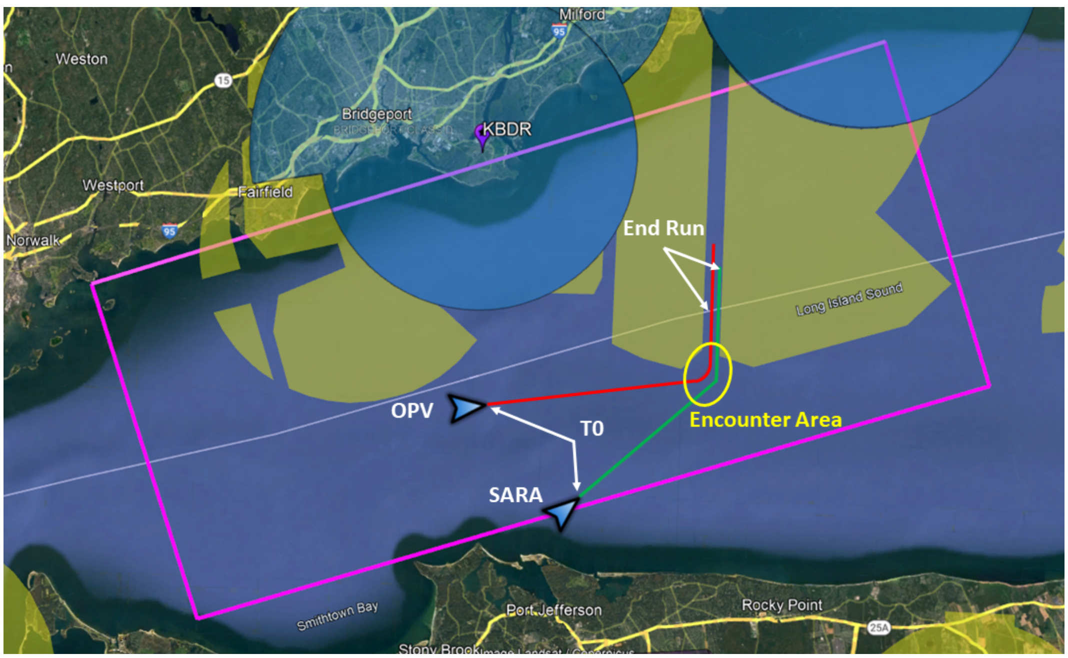


Figure B6. Maneuver 39 Merging Geometry Over Long Island Sound Airspace (0 Winds)

Appendix C: Flight Test Metric Results

C.1 Metric 1: False Detection Rate

Description: For all conflicts without executed resolutions, compare first detection conflict time and predicted CPA separation distance with actual time and CPAs to determine if a conflict actually existed.
Analysis: Across 1082 CD cycles from 34 conflicts, there was a false positive detection rate of 0.92% (per cycle) and 2.94% (per conflict). Maneuver M20 contained the false positive detections, which were with OPV.

C.2 Metric 2: Missed Detection Rate

Description: For all encounters between the ownship and traffic, assess actual CPA distance to determine if a conflict actually existed.
Analysis: Across 41 runs and all CD cycles, there was a false negative detection rate of 0%. All LOS events were successfully detected by AOP CD capability.

C.3 Metric 3: Actual vs. Predicted CPA Error

Description: For all executed resolutions, compute the average deviations between the predicted CPA of resolution trajectories and actual achieved CPA.
Analysis: All resolution maneuvers achieved separation. To fairly compare the predicted and actual CPA, runs that ended before the point in time when the predicted CPA occurred are excluded. Given that, 15 runs were retained. Furthermore, the deviation is quantified in the direction that the CR occurred. For the lateral and speed resolutions that were retained (M1, M4, M6, M19, M20, M22, M32, M42), the average lateral deviation between the predicted and actual CPA was 142.5 feet. For the vertical and hybrid resolutions that were retained (M2, M8, M21, M23, M31, M41, M43), the average vertical deviation between predicted and actual CPA was 15.6 feet. Results are shown in Figure C1 for each maneuver.

C.4 Metric 4: Conflict Detection Alert Delay

Description: For all conflicts, compute the time difference between the initial conflict alert and the triggering event. A triggering event may occur at the CD look ahead horizon or be caused by an external trigger.
Analysis: Across 32 runs with Conflicts with Intruder, all conflicts were detected within three seconds of the triggering event. Triggering events were normal CD horizon cases except:

- M31, M32: OPV intent change caused conflict.
- M40: SARA intent change caused conflict.

A distribution of conflict alert times for all runs is shown in Figure C2.

C.5 Metric 5: Rate of Resolvable Conflicts

Description: For all conflicts, determine the rate of occurrence for which at least one resolution was provided before end of run (typically 30 seconds prior to LOS or a "knock-it-off" call).

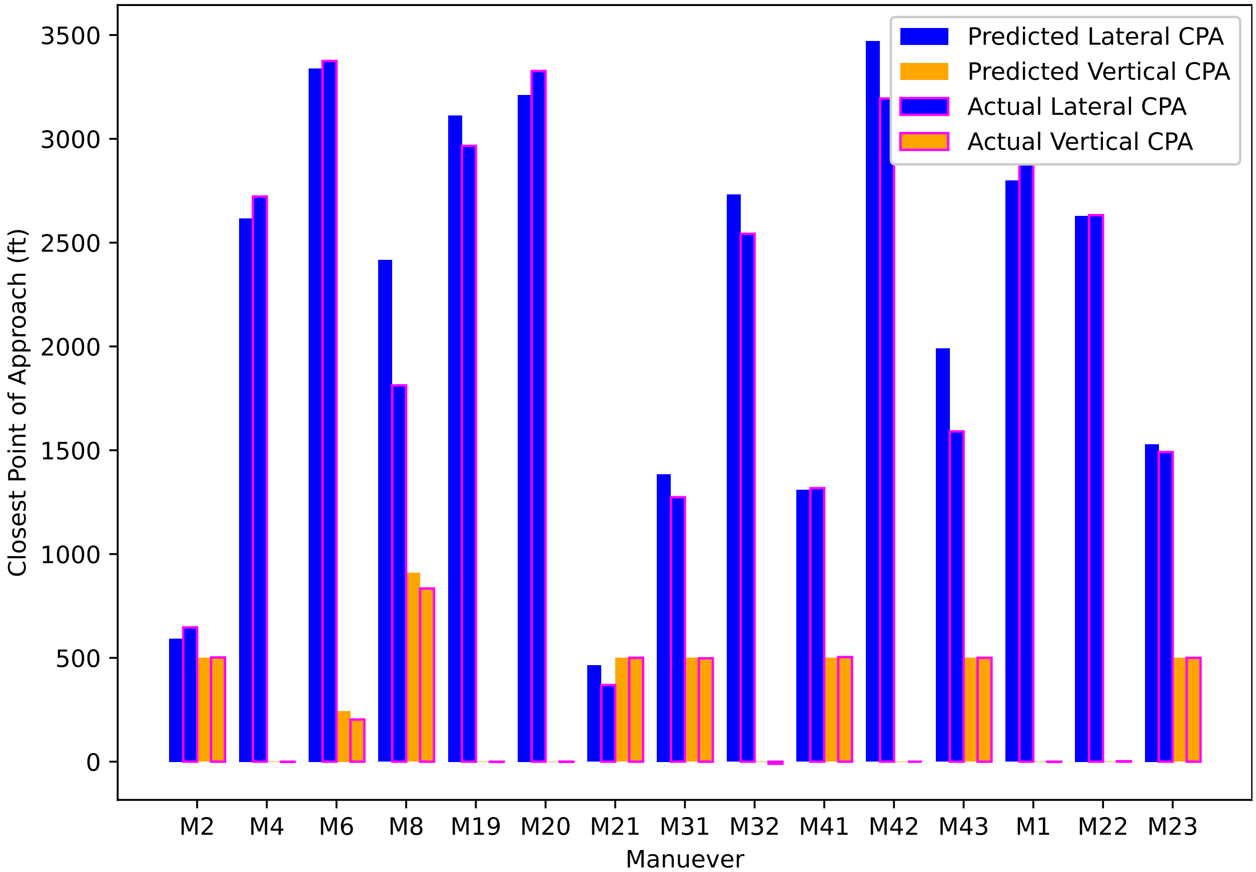


Figure C1. Flight Test Metric 3 Results. CPA in maneuvers with resolution only.

Analysis: Across 32 runs and 33 conflicts, 97% of conflicts had a resolution provided before T30. Maneuver M40 experienced a conflict that could not be resolved.

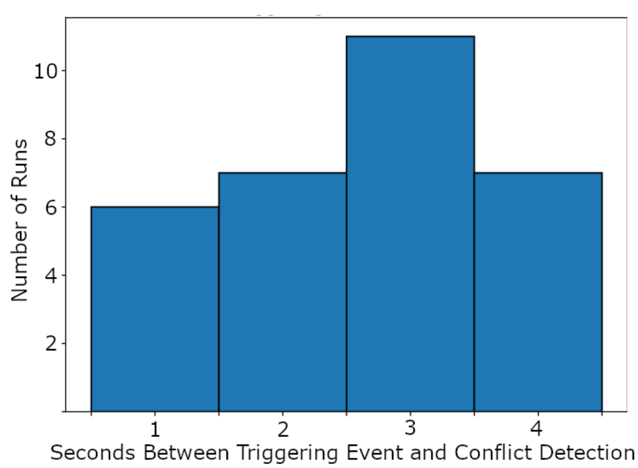


Figure C2. Flight Test Metric 4 Results. Histogram of time between triggering event and conflict alert.

C.6 Metric 6: Same-Traffic Conflict Occurrence Rate

Description: For all conflicts with executed resolutions, determine if a second conflict occurred with the Intruder traffic from the previous conflict.
Analysis: Across 23 conflicts, each having one executed traffic CR, no runs had a repeat conflict with the same intruder. Maneuver M39, a corridor maneuver, predicted a second conflict, with LOS occurring near OPV's vertiport landing, but the conflict's predicted first LOS was beyond the ownship's CR look-ahead horizon.

C.7 Metric 7: Actual LOS Rate after CR

Description: For all conflicts with executed resolutions, determine the percentage of LOS that occurred when flown (whether AOP detected it or not) via observed state.
Analysis: 0% of executed CRs had a LOS (based on observed aircraft positions). Lateral and vertical separation distances at CPA for each maneuver maintained minimum separation criteria (1500 ft lateral and 450 ft vertical).

C.8 Metric 8: Rate of First CR Cycle Resolution for All Conflicts

Description: For all conflicts, calculate the rate of occurrence where at least one resolution was provided within the first CR cycle.
Analysis: Across 34 traffic conflicts, 97% received at least one resolution solution from AOP within the first CR cycle. Maneuver M40 experienced a conflict that could not be resolved.

C.9 Metric 9: Rate of First CR Cycle Resolution for Resolvable Conflicts

Description: For all conflicts where at least one resolution is found, calculate the rate of occurrence where at least one resolution is found within the first CR cycle.
Analysis: Across 33 resolvable traffic conflicts, 100% had a resolution within the first CR cycle.

C.10 Metric 10: RTA Conformance Rate for Traffic Conflict Resolutions

Description: For all executed CRs (triggered by traffic conflicts), determine the fraction of RTA conformance magnitudes that are within the maximum acceptable time tolerance.
Analysis: For runs with a conflict that executed a resolution, 90% had a final estimated time of arrival (ETA) within the acceptable time tolerance. M17 is excluded because it was intentionally designed to not be able to meet the RTA. M40 is excluded because no resolution advisories were generated. Figure C3 shows additional information. ETAs are shown for each maneuver. Positive values correspond to late arrivals. ETAs resulting from PBGA executions are shown by red bars. ETAs resulting after execution of the initial speed advisory are shown by blue bars for M15, M16, M17, and M40, which were meet-time maneuvers designed to also create a conflict. For these maneuvers, a ETA outside of RTA tolerance triggered a PBGA execution. Maneuver M40 shows no PBGA solution.

Figure C4 displays a distribution of RTA conformance error for all 280 AOP advisories, including maneuvers having RTA changes that did not involve a traffic conflict. 35 advisories were outside the three second tolerance.

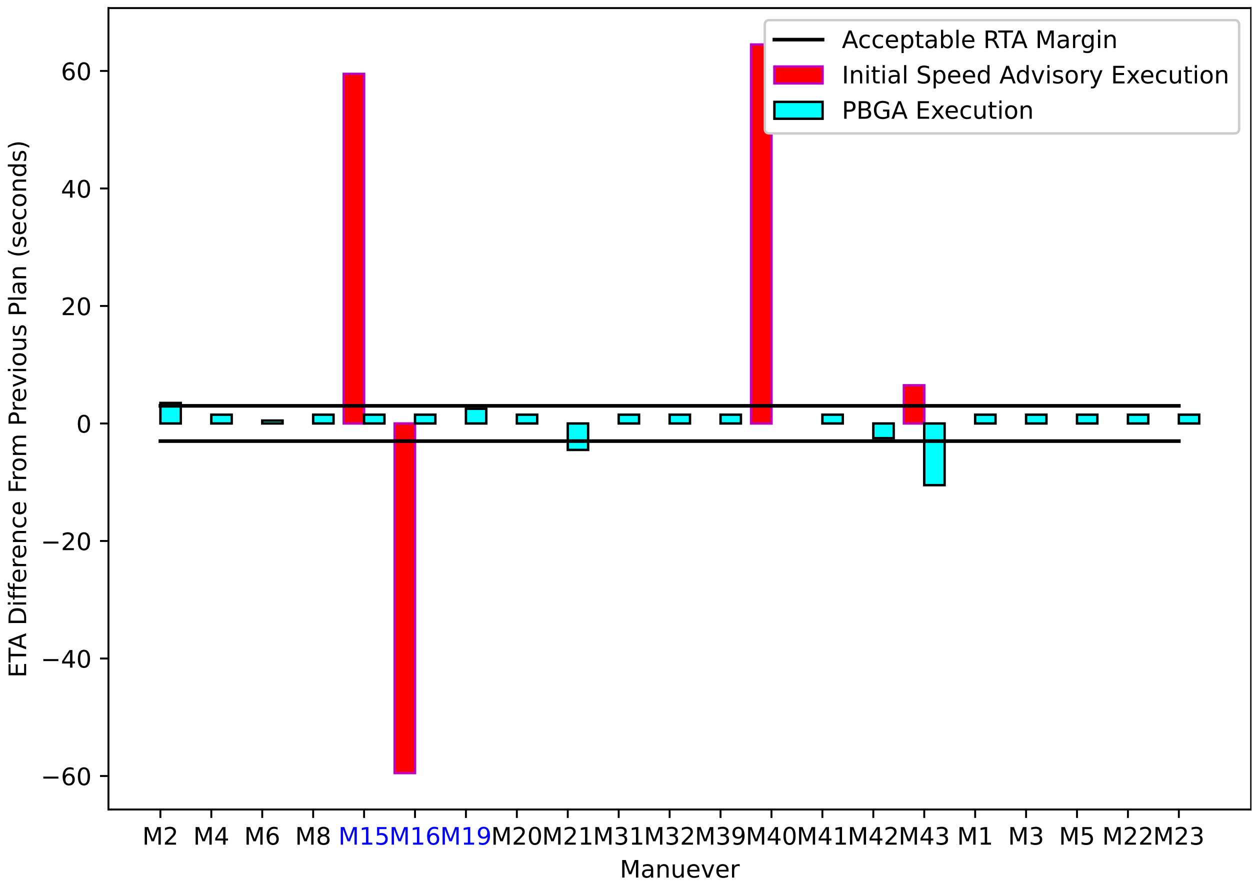


Figure C3. Flight Test Metric 10 Results. ETA margin for executed runs only.

C.11 Metric 11: Conflict Prevention Rate

Description: For all conflicts with executed resolutions, compute the occurrence rate of future conflicts with any traffic caused by the resolution of an existing conflict. Future conflicts outside the CR horizon at the time of CR computation are not considered by the metric.
Analysis: Across 21 conflicts with executed resolutions, 0% of all executed CRs resulted in a future conflict within the CR horizon. M39 contained a second conflict with OPV, but that conflict occurred beyond the CR horizon.

C.12 Metric 12: RTA Conformance with RTA Change Requests and Traffic Conflicts

Description: For all conflicts with executed resolutions, compute the average magnitude of RTA deviation caused by an RTA change request and a traffic CR.
Analysis: Across two runs, 100% of all executed CRs were within the acceptable time tolerance. The average deviation was one second. Maneuver M17 was excluded because it was designed to not meet the RTA. For more detail, see Figure C4. The two runs are M15 and M16.

C.13 Metric 13: Actual vs. Predicted CPA Error in Corridor Operations

Description: For all conflicts involving corridor operations with executed resolutions, compute the difference between actual vs. predicted CPA after CR.

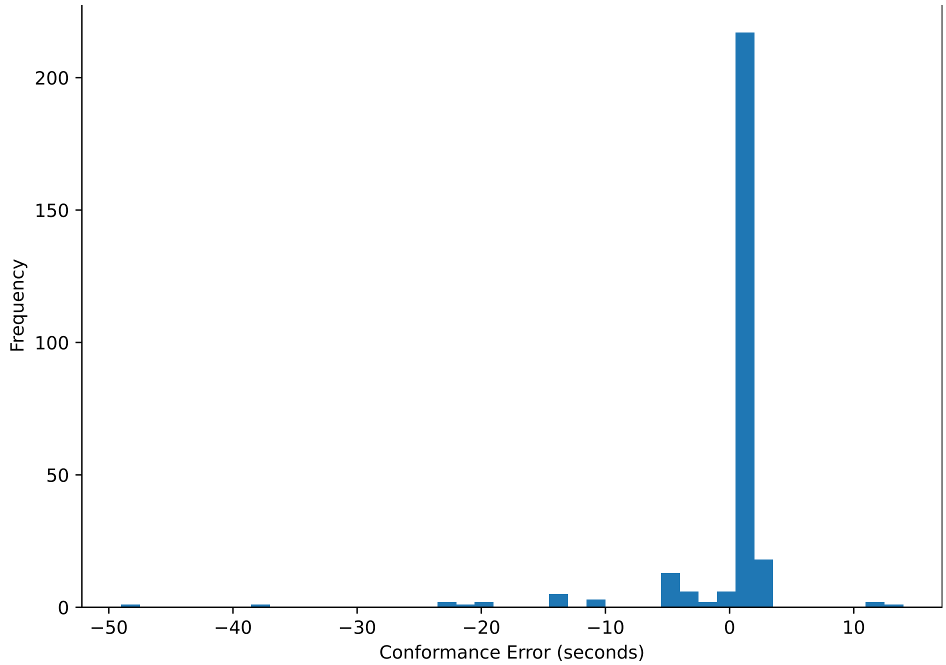


Figure C4. Flight Test Metric 10 All Advisory Results. RTA conformance for all strategic intent-based conflict resolution (SICR) advisories. N = 280.

Analysis: Across three runs, the mean lateral deviation between actual and predicted CPA was 730 ft and vertical deviation was 120 ft.

C.14 Metric 14: Conflict Prevention Rate in Corridor Operations

Description: For all conflicts involving corridor operations, a) compute the occurrence rate of future conflicts with any traffic caused by the resolution of an existing conflict, and b) determine the rate of resolvable conflicts involving corridor operations. Future conflicts outside the CR horizon at the time of CR computation are not considered by the metric.
Analysis: Across three runs having two executed resolutions,

- 0% of executed CRs resulted in a future conflict within the CR horizon.
- Out of 4 conflicts involving corridor operations, 75% were resolvable.

Maneuver M39 contained a second conflict with OPV, but that conflict occurred beyond the CR horizon. Maneuver M40 did not produce a resolution.

C.15 Metric 15: CR Performance Evaluation

Description: Computed the elapsed time of all PBGA resolution advisories and determine the average elapsed time for the entire data set and for each resolution degree of freedom.
Analysis: The analysis set was made up of 213 CR cycles across 33 runs. Run M10a was omitted due to its abnormally high count of sub-second CR processing cycles.

- The average total processing time for PBGA CR is 5.41 seconds.
- Average population processing time in PBGA CR:

- Lateral: 4.55 seconds
- Vertical: 2.74 seconds
- Speed: 3.87 seconds
- Hybrid: 5.01 seconds

C.16 Metric 16: CR Trajectory Flyability Evaluation

Description: For executed CR trajectories, determine whether any AOP ownship performance model limit was exceeded by any portion of the trajectory. For the flight test, performance model constraints consist of set of speed constraints and climb or descent rates.
Analysis: All resolution trajectories were flyable by the ownship.

C.17 Metric 17: CR Trajectory Performance Limits Evaluation

Description: For executed CR trajectories, determine whether any portion the CR trajectory was at a limit specified by the AOP ownship performance model. For the flight test, performance model constraints consist of set of speed constraints and climb or descent rates.
Analysis: Many resolution trajectories required the ownship to operate at a limit. Details are provided in the body of the report.

C.18 Metric 18: CR Cycle Elapsed Resolution Display Time

Description: For all conflicts, compute the average elapsed time between consecutive completed PBGA CR cycles. The PBGA completion event time is an estimate of the time pilot operators would first see a cycle's resolutions presented to an external display interface. The difference between two PBGA completion event times is an estimate of the time the operator has to evaluate, select, and execute resolutions.
Analysis: Across 33 runs having 124 cycles with presented resolutions,

- The average time between displayed resolution sets is 17.86 seconds.
- 118 sets of resolutions were displayed for 15 seconds or more (95.2% of all resolutions).

Display durations are bounded by the PBGA computation end time of the current CR cycle and the PBGA computation end time of the next CR cycle. Data points were excluded for the following cases: a) the cycle associated with a pilot's resolution execution, which prevents completion of the subsequent CR cycle, and b) the end of a run, thereby preventing completion of the subsequent cycle.

Figure C5 provides distributions of display time durations. The lowest two time durations were 6.13 seconds (Run M1, Group 2) and 9.5 seconds (Run M4, Group 1). The baseline CR refresh cycle time was 20 seconds.

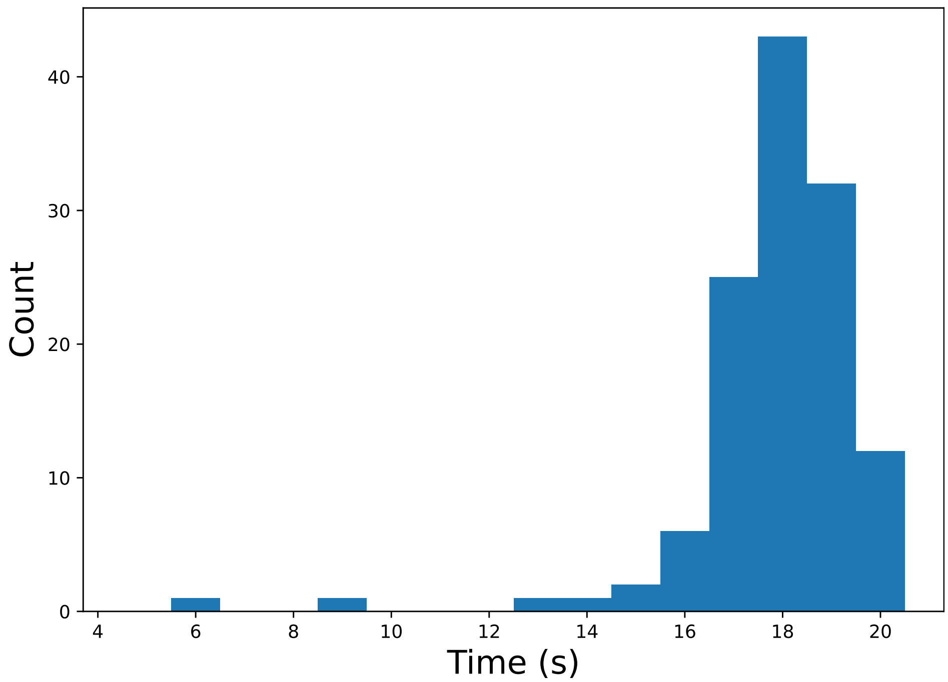


Figure C5. Flight Test Metric 10 all advisory Results. Times between PBGA CR cycle completion.

Appendix D: Research Pilot Interview Responses

The following is a transcription and condensation of research pilot interviews conducted at the conclusion of the test. Interviews were conducted with each research pilot separately. The transcription has been reviewed by the two research pilots and includes their updates.

D.1 Written FPM Post-Test Questions

1. I understand AOP's functions and behaviors. (1: strongly disagree; 2: disagree; 3: neither agree nor disagree; 4: agree; 5: strongly agree)

   Pilot 1:

   > Agree (4)

   Pilot 2:

   > Agree (4)

2. I was never confused by AOP's alerts and advisories during flight. (1: strongly disagree; 2: disagree; 3: neither agree nor disagree; 4: agree; 5: strongly agree)

   Pilot 1:

   > (Pilot 1 also addressed several other questions here.) Neither agree nor disagree (3). On the surface level, what we're doing is easy. But one level down, i.e., four resolution options to choose from, it gets harder. Lateral is easy. The other options are harder. The speed option was always confusing. (Pilot 1 talked about the display interface, concentrating on the profile display.) Knowing intruder's altitude would probably be helpful. (Comment: relative altitude should have been displayed.) Pilot 1 also felt the task was easier to perform in simulation than in flight. He also mentioned there's not enough time to evaluate the resolution advisories. He implied that even two minutes might not be enough. He approached the task with the attitude, "this one's good enough."
   >
   > Pilot 1 noticed the "random" alerts and resolutions. We mentioned we also noticed that and didn't understand what was happening. (We now believe this was a result of the winds information not getting into AOP, so AOP was assuming zero winds and was sometimes having trouble with the RTA meet-times.)
   >
   > Question: You said 20 seconds of the refresh cycle was not enough to evaluate resolution options. How did you feel about the 40 seconds case [which was part of Test Group 5]? Pilot 1: 40 seconds is probably enough time. Question: What if AOP provides the best solution? Pilot 1: If AOP advises on which resolution is the best, that solves that [i.e., 20 seconds would be enough]. But we also need to know when the conflict comes up. We're looking at it [in this test], but otherwise we would need an audio alert.

   Pilot 2:

   > Neither agree nor disagree (3). No, not confused. Started as a yes [I was confused], but with a little familiarity with the system, this turned into a no.

Pilot 2 said he needs to see the time-to-separation-loss indication better. That should be well positioned, large, and clear.

Instead of the conflict indication and intruder intent disappearing after a resolution is selected and the conflict is resolved, Pilot 2 would like to see the conflict aircraft highlighted so he can monitor it. A researcher mentioned previous displays where you could click on an aircraft to get state and intent information. Pilot 2 said that would be useful.

Question: How does Pilot 2 feel about the "cotton ball" indicator? Pilot 2: It's good, but it's kind of a new concept. There's nothing like it on any display. He needs to develop more familiarity with it. "I tend to look at the raw crossing paths." Question: (A researcher described the previously used "dog bone" indication.) Would something like that be better? Pilot 2: Yes, I tend to think about what's on my path. [The dog bone indication] is more intuitive and less distracting than the circle. I would also like to see the trajectory of the intruder.

3. I never felt a need for additional information to understand what AOP was advising. (1: strongly disagree; 2: disagree; 3: neither agree nor disagree; 4: agree; 5: strongly agree)

   Pilot 1:

   Disagree (2). There were meet time issues, difficulty in figuring out speed resolutions, and vertical resolutions difficulty with the "side view." The hybrids were often very similar to the vertical ones. Pilot 1 agreed with the statement that most of his issues could be resolved with a better UI. He introduced an idea for AOP to withhold some resolution advisories unless there is a reason to show them. Reasons to show include they're significantly different from the others, or if they are more optimal than the others.

   Pilot 1 observed a hybrid solution case where the ownship climbed while the intruder was already at a higher altitude. The intruder may have been planning to descend. Pilot 1 said for this and some other cases, AOP resolutions were counterintuitive.

   Pilot 2:

   Disagree (2). Need additional information on the display to know things like when a climb will start. There was too much information (such as diamonds on display) to be clear to pilot so he can quickly make decisions. "Sometimes that CPA I don't necessarily comprehend where it fits into my timeline." Needs something that gives better situational awareness about time and distance, especially for the profile display. More context in the display, such as turn points on the profile display, would help with situational awareness (SA).

4. All presented trajectory resolutions seemed reasonable for use in routine operations. (1: strongly disagree; 2: disagree; 3: neither agree nor disagree; 4: agree; 5: strongly agree)

   Pilot 1:

   Agree (4). All climbs, descents, and speed changes were in the normal range. All solutions were well within Pilot 1's understanding of the aircraft's performance.

   Pilot 1 generally assumed the system had more information than he did, and would produce adequate resolutions, although he did mention trust would be a factor. Pilot 1 also said, "I never saw one [where I thought] oh my god, I would never want to do that."

In response to a question, Pilot 1 said he executed the hybrid resolution most often, mostly out of curiosity.

Question: Did you ever see a case where you questioned the resolution, such as "why is it taking me so close to that aircraft?" Pilot 1: Yes, a few times. The classic one for me was when they're coming at us, with a lateral solution, it seemed like it would turn us back before we were abeam of the intruder.

Question: One thing we do now is treat the protected zone as a hockey puck shape. How would you feel about a different shape, such as an ellipsoid? Pilot 1: It makes sense to have more separation if you're going in front of someone, but I would think that is somewhat built in. Intellectually, it doesn't seem relevant, but in practice, larger buffers in front of the conflict may feel more comfortable.

Pilot 2:

Agree (4). "All the resolutions seemed reasonable, but I wondered why there were such variable nuances, and why in one group of resolutions, it's a quick turn back to the route and in the next set of resolutions, it's a turn and a drive of many many many miles away from the turn and then back on to the route, and I'm like, why? If there are other aircraft or airspace factors that are involved in the decision, I feel those should be obviously highlighted when we preview that." It helps to build SA. Pilot 2 said it would also help to identify which resolution AOP would recommend as the most efficient. One idea that Pilot 2 liked was to have AOP present only its preferred solution. If the pilot doesn't like it, he can reject it, thereby allowing AOP to present another solution.

Pilot 2 said lateral resolutions are his preferred option. He also said he would generally never choose to cross the flight path of a conflicting aircraft to get to the other side of it. (It's possible there's other traffic causing that, but with our current FPM Engineering UI, the pilot doesn't have that information.)

Question: Would a different shape for a protected zone be better, such as an ellipsoid instead of a hockey puck? Pilot 2: Yes, I support that. If I know I'm going behind you, I can cut it close.

5. Were you less comfortable operating with AOP for certain test runs than others? If so, which test runs and why?

Pilot 1:

No, I was always comfortable with it. Sometimes I didn't understand why it did something, but I was not uncomfortable.

Question: Any thoughts about operating in corridors? Pilot 1: I don't think corridors are necessarily a problem, but you need to have your act together. Weather could completely explode that because you don't have options. On a clear blue day with everybody behaving, I don't think it would be [difficult], unless you have a lot of traffic.

Pilot 2:

No strong opinion. Corridors make Pilot 2 "feel more constrained than anywhere else," and wonders whether he will get quality resolutions that work in the constraints of a corridor. (Pilot 2 asked if some compensation was used or could be used. He may have meant something like increasing the look-ahead horizons in corridors, or he may have simply been asking if more sophistication in the algorithm is needed to handle corridors.)

General comment: Pilot 2 would like to see the system stressed where changes happen continuously, to see how it handles "a really bad day," to build trust.

6. Which resolution types did you tend to feel least comfortable with? Why?

   Pilot 1:

   > "Speed seemed harder to evaluate to me. [But] I don't have any reason to think [the resolution] would be bad either." The hybrids seem to be mostly vertical with just a sliver of lateral. He was wondering "why is [hybrid] even there? But that's not really a comfort thing."
   >
   > Question: Given the four solutions, were there ones that seemed more intuitive than the others in terms of "I know this will solve the problem"? Pilot 1: Yes, for me, from the perspective of understanding of what it's going to do, the vertical solution was. If you've got a problem with someone, just get above them until you're past them. But from a pushing-the-buttons perspective, the laterals were the most obvious. If I had more experience with the system, [the others] would be easier. The speed solution was very difficult with the experience level I had with the display. Really look at that side view and figure out what it needs to do from a speed standpoint. Intuitively, speed should be okay.

   Pilot 2:

   > "I'm a little uncomfortable with all of them, but because I didn't have good awareness of the vertical view, the hybrids and verticals make it a little more difficult to know exactly when and for how long that maneuver is going to be active. Maybe I'm about to start an approach soon or maybe I'm about to change my flight plan shortly. There could be factors that could play in to [not knowing how long the maneuver will be active]."

7. What operational situations or flight factors would influence your prioritization of available options in resolving conflicts and meeting RTAs?

   Pilot 1:

   > "I think the one most pilots would consider is, is this going to cost me more energy? Fuel, battery, hydrogen, or whatever we're going to be using in these things." He also mentioned if he is in the clear and a vertical resolution would put him in clouds, he wouldn't select it. He also wouldn't head into a storm.
   >
   > Question: If you had passengers in the back, would that change any of your decisions? Pilot 1: I would think no, but maybe if some resolutions got closer [to hazards] and might look scarier, I might not choose them. When asked about passenger comfort, Pilot 1 stated that none of the maneuvers are uncomfortable except possibly speed-ups and slowdowns, [so speed could be a factor in implementation,] depending on how aggressively the system does those maneuvers.

   Pilot 2:

   > This is very situation-dependent. Proximity to a very busy terminal area [is very different from when] I'm in en route cruise. All these factors influence why I would choose one over the other.

If they're all very similar, I would base it on personal preference — I don't want to climb, or I just want to turn. Changing altitude will burn more fuel, so most folks are going to want to make lateral changes. Climbing burns fuel. Descending doesn't burn fuel, but then staying at the lower altitude burns fuel. Pilot 2 also implied that you're more likely to encounter turbulence if you change altitudes than if you stay at your altitude.

8. What additional information would you like AOP to provide, if any, to help you prioritize and select from the available resolution options?

   Pilot 1:

   > "If there's a way to [have AOP] more clearly let me know what its plan is, I think that would be better information." Cost and comfort are the two things Pilot 1 mentioned as understanding the "why" of a recommendation. At the end, he also mentioned "if it knew one [advisory] was taking you closer to hazardous weather, that would be good to have in there." When prompted, Pilot 1 also said if everything else was equal, he would prefer more separation from a conflict if the additional separation was not excessive. He used 10 miles as an example of excessive separation.

   Pilot 2:

   > (We already talked about this.) Make the conflict aircraft's intent either selectable or still showing after execution.

9. Based on your experiences with AOP, do you have any suggestions for improving the trajectory solutions AOP provides?

   Pilot 1:

   > Pilot 1 again brought up the hybrid solution as being odd. Hybrids were typically mostly vertical resolutions with very small lateral deviations.
   >
   > Question: Would you find it more unnerving to resolve conflicts within a corridor vs. resolving before or after a corridor? Pilot 1: I think there would be no difference, provided it can be resolved. The corridors will need to be designed to allow for resolutions within them; corridors "need to be spaced to resolve" conflicts. Perhaps a vertical passing lane.

   Pilot 2:

   > Question not asked.

D.2 Verbal FPM Post-Test Questions

1. What are your thoughts about the separation standards we assumed for UML-4 and used in this test? (1500 ft laterally, 450 ft vertically)

   Pilot 1:

   > Pilot 1 mentioned that in a visual environment, our quarter-mile lateral separations seemed large enough. He said that may not apply to an Instrument Flight Rules environment. (Comment: Pilot 1 is talking about VMC vs. IMC. We have been assuming the UML-4 operations will not rely on visual separation.)

Question: What do you think of the sizes of buffers around the protected zones? Pilot 1: In general, having more there would be good, but knowing what we're trying to get to with this [concept], maybe not. Maybe pilots will get more comfortable with the higher traffic density. It didn't really seem like, 'Oh my gosh, that's way too close!'

Pilot 2:

I think [separation standards] could be scaled down for the urban environment with electric vertical take-off and landing (eVTOL) (including a mixture of full-size aircraft and small unmanned aircraft) if we're in a structured airspace with lots of corridors, or even if we have onboard algorithms making real-time deconfliction decisions. I would see that intent-sharing would be important and it would need to be standardized and updated in real time. It could be overwhelming, so we need to have a way to deal with that possibility. If intent keeps changing because all the airplanes are interacting with each other, and each aircraft changes its intent to make its situation better, it's a ripple effect. [Reminds me of] chaos theory. If we take a slow build up approach to determine the point of saturation, and then have the system keep things under the saturation level, [it could work].

2. The traffic you experienced is modeled to be representative of a future UML-4 operational environment. Do you have any thoughts about operating routinely in such an environment?

   Pilot 1:

   You need to experience it to really know. "If you look out and see six dudes out there, coming every which way, [you] hope everybody knows what they're doing." Pilot 1 also mentioned that if we were up and away, the traffic density during the test would be concerning, but down low, you're going much slower, it's Visual Flight Rules, you can see where they are and see where they're going, so there are less worries.

   Pilot 2:

   I could operate routinely in a UML-4 environment. It didn't seem out of the ordinary. Just the intent part is new. I need to get used to operating with a bunch of intent. It could be everyone's intent is just too much to comprehend. At one point I looked at [another part of the overlaid DFW environment] and there was a continual sea of traffic symbols, all overlapped. Give me just the information I need to know when I need to know it.

3. Do you have thoughts about whether UML-4 operations could safely be conducted in all meteorological conditions?

   Pilot 1:

   "If we had systems with low technical errors and deconfliction systems like this, I wouldn't think there would be a problem." Pilot 1 added his comments don't include flying in hazardous weather.

   Pilot 2:

I think this could be easily done in Instrument Flight Rules if the system is robust to show thunderstorms or other weather hazards. (We told Pilot 2 we already have that, and AOP will route around them, but a test scenario covering it just didn't make the cut.)

4. Do you have any recommendations for what is needed in future testing?

   Pilot 1:

   More traffic in the scenario. The display needs to be more visible; better integrated (e.g., part of scan, not on a tablet). Also develop a pilot display instead of an engineering UI. Pilot 1 agreed it would be important to equip an intruder aircraft with AOP. He also mentioned exploring [along-path] time conformance requirements. "How tight is tight enough? Industry likely won't have this, and will ask why we need such tight time conformance."

   Pilot 2:

   Question not asked.

5. Additional time horizon questions

   - Under any of the time parameters variation runs, did you feel you had more time than necessary? For instance, did AOP alert you to a conflict earlier than you felt necessary?

     Pilot 1:

     Question not asked.

     Pilot 2:

     Pilot 2 didn't notice any difference in time horizons. He expected to be able to tell a difference, but he couldn't.

     Regarding time available to look at resolutions, Pilot 2 said he was generally hurrying because he didn't have a sense of how much time he had remaining to make a selection. He said the buttons don't blink. The "new resolutions" banner would be an indication, but on the tablet, it was very hard to see. If there were two or three resolution options, he needed "two or three times the amount of time to make a decision. If I just had one resolution, I had plenty of time. I probably need five seconds for that."

   - How did you feel about the conflict alert look-ahead times? Our default was three minutes.

     Pilot 1:

     I didn't see a difference [when we varied the look-ahead times]. That's more a function of how hard it's going to be to resolve the conflict, right? From a comfort standpoint, if AOP has enough time to take me around the conflict, [I'm comfortable]. Pilot 1 also suggested it might be useful to analyze pilot responses in today's Visual Flight Rules operations to determine how far ahead in time they are looking.

     Pilot 2:

     Question not asked.

- Any specific comments about our Group 6 maneuvers, where the intruder changed its intent and created a conflict at 2 minutes and 1 minute to first LOS?

  Pilot 1:

  Pilot 1 didn't notice a difference between these time horizons and the default time horizons.

  Pilot 2:

  Pilot 1 performed the two Group 6 runs, so Pilot 2 doesn't have opinions about them.

6. The only SA you got for virtual elements (background traffic, SUA) was through the FPM engineering UI. Do you believe that SA was compelling enough to impact your decision-making during a run?

   Pilot 1:

   Question: Do you feel like you were making decisions in flight as if the virtual traffic was real? Pilot 1: In general, I was not concerned about the virtual traffic, primarily because I assumed AOP would handle deconfliction from them. In order to [help give me the feeling of the virtual traffic being a real threat], it might help to have some conflicts with virtual traffic in test scenarios.

   Pilot 2:

   I was concentrating on my task during the test (such as looking for OPV to show up), and with the FPM Engineering UI I didn't have a way of concentrating on the aircraft important to me (such as within 2000 feet of my altitude), so [decision making] probably would be different if they were real airplanes. Even if you put me in a [simulator with realistic out-the-window views] (unintelligible), [decision making probably would be different].

Appendix E: Measurements of Conformance to 4D Trajectories

As mentioned in Section 8.1.1, conformance of an aircraft to its 4D trajectory was measured by the FPM/AOP TPUBs Calibration Analysis Tool (FATCAT) (Ref. 10). The input for FATCAT was prepared by collecting the applicable set of EUTL plans that were used for each aircraft's guidance, determining the time interval during which each of these plans was active (from when it was first executed to when another EUTL plan was executed or the run ended), and collecting the sequences of recorded aircraft latitude, longitude, and altitude with timestamps (the "observed states") during those time intervals. In some test points, either SARA or OPV (or both aircraft) ended their runs before the end of test was called and (in the case of SARA) before terminating AOP. The times at which these events occurred were estimated for each aircraft and observed states recorded after those times were excluded from the analysis.

E.1 FATCAT Algorithm

In order to measure the deviations of observed states from the 4D flight plan, FATCAT partitions the plan into a sequence of straight segments (geodetic paths between waypoints) and turn segments (paths that follow small-circle arcs). FATCAT maps each observed state onto one of these segments. A state mapped onto a straight segment is projected orthogonally onto the segment as shown in Figure E1a. A state mapped onto a turn segment is projected radially onto the segment as shown in Figure E1b, using the center of the turn arc as the center of the projection. (The deviations of the observed states from the EUTL plan in Figure E1 are exaggerated for the purpose of illustration.) The point to which the observed state is projected is its abeam point.

The altitude deviation of an observed state is the signed difference of the altitude of the observed state and the predicted altitude of the EUTL plan at the abeam point. The time deviation is the signed difference of the recorded time of the observed state and the predicted time at the abeam point according to the EUTL plan. The magnitude of the cross-track deviation is the distance to the abeam point; FATCAT distinguishes between deviations to the left and deviations to the right of the path by the sign of the deviation and makes a separate record of deviations from turn arcs in which the sign indicates whether the observed state is outside the turn arc or inside it.

E.2 Results for SARA

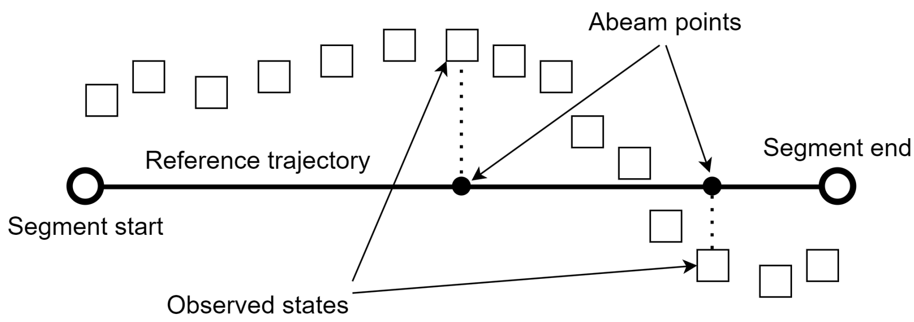
SARA's along-path (time) and cross-track (lateral) conformance were excellent during most of the flight test runs. In almost all runs, the ETAs of all observed states occurred well within tolerance at the abeam point. Cross-track errors were well within tolerance in almost all runs. Altitude errors were almost all acceptable (within ± 25 feet) as well.

A few exceptions to the conformance were observed, as listed below.

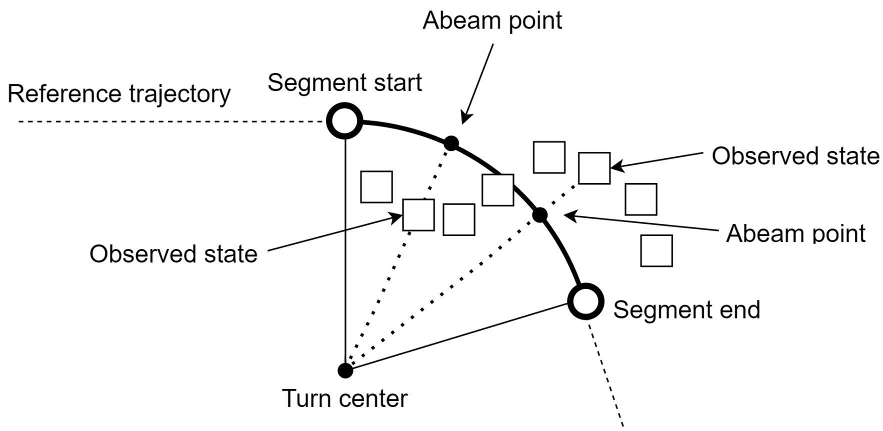
E.2.1 Along-Path Errors

In the run for M13b, time errors ranged from -5.10 seconds to 3.58 seconds. This run included a speed change to meet a new RTA after T_0, followed by a CR that lowered the speed by about 13 knots.

In M20, the time errors ranged from -0.34 second (in conformance) to 3.64 seconds. This run included a speed-only maneuver to resolve a conflict, for which Behavior-Based Trajectory Generator (BBTG) mistakenly predicted an immediate speed increase at a point where the maneuver pattern called for it to hold the previous airspeed.



(a) On a straight segment.



(b) In a turn segment.

Figure E1. Abeam points used in measurements by FATCAT.

E.2.2 Cross-Track Errors

In M32, SARA was up to 432.5 feet to the left of the planned flight path. This was attributed to a lateral CR that the research pilot executed after the aircraft had already passed the point where the turn was intended to begin. Remarkably, SARA maintained along-path (time) conformance during this run.

E.2.3 Altitude Errors

In M8, which executed a vertical CR, SARA was 44.6 feet below the predicted altitude at one point during the climbing maneuver. It appears that the resolution flight plan was executed at almost the same instant that the climb was planned to begin.

In M16, which executed a hybrid lateral-vertical CR, SARA was 88.6 feet below the predicted altitude at one point during the climbing maneuver. There was no apparent reason for the delay in the climb, since the new flight plan was executed several seconds before the climb was planned to begin.

In M23, SARA was briefly up to 27.8 feet above the predicted altitude from 11 seconds to 13 seconds after T_0, that is, immediately after the research start time.

In M43, after executing a resolution, SARA was 27.3 feet below the predicted altitude near the end of a planned climb to the new cruising altitude.

E.3 Results for OPV

Due to questions about the time conformance of OPV, the state data for the OPV trajectory conformance for this report were taken from cvtLog files rather than from the data collected by AOP aboard SARA. After the analysis, however, spot checks of several runs confirmed that the data collected by AOP agreed exactly with the data from the cvtLog files.

OPV appears to have ended a run early more frequently than SARA. Runs that OPV ended early included M1 (group 1), M3 (group 1), M22, M42, and M43.

With a few exceptions (which are described below), the observed states for OPV were all well within bounds. In several runs, however, the recorded times consistently averaged one or two seconds “late.” The differences between actual and predicted times tended to vary less for these runs than for many of the SARA runs; there was simply a small systematic bias in the observed times whose cause was not determined.

E.3.1 Altitude Errors

When running M21, OPV started out of conformance with the 4D flight plan (about 7 seconds late and 257 feet left of the intended track), but it regained lateral (cross-track) conformance within 4 seconds from T_0 and along-path (time) conformance within 29 seconds from T_0. This appears to have been an error in the procedure for starting the run at T_0 rather than a failure of the guidance system to maintain conformance. Later in the same run, however, the altitude varied up to 26.1 feet below and up to 46.6 feet above the predicted altitude.

E.3.2 Possible Loss of Precision in Recorded Data

In the run for M19, many of the OPV observed states seemed to suffer from loss of precision. States with suspected precision loss had altitudes that were exact multiples of 25 feet. Several of these states had latitudes and longitudes that indicated that the aircraft had flown backward from the previous observed state.

FATCAT measured violations of along-path (time) conformance for several observed states in this run. All violations showed the aircraft more than three seconds late. All of the observed states with violations showed evidence of loss of precision. The “backward” states had the largest measured along-path violations of the run, up to eight seconds late.

When the states with suspected loss of precision were eliminated, however, all remaining observed states (including states that were technically in conformance) were still at least 2.4 seconds late. This suggests that time was not completely synchronized across all components in this run.

Appendix F: Effect of Missing Wind Predictions on Test Runs

As explained in Section 8.1.2, during the flight test, due to errors in development and integration, AOP used zero wind speed rather than the predicted wind in its calculations of flight plans and resolutions.

The following is a detailed analysis of the test runs indicating how they were determined to have been affected or not affected by AOP's use of zero wind speed. The test runs are listed in the sequence in which they were run, organized by date.

F.1 October 18

On October 18, the predicted wind was zero, so the wind used in AOP was accurate. No runs that day were affected by AOP's use of zero wind speed.

F.2 October 19

On October 19, the predicted wind was 10 knots from 180 degrees. This corresponded to just a small headwind on any of SARA's headings. The maximum cruise CAS used in AOP's active routes was about 101 kt CAS, observed in maneuver M17. The headwind effect was almost certainly less than 8 knots. No other run was even close to the upper speed limit.

F.2.1 Maneuver M17

The last FPM run of the day, M17 with 10 kt wind from 180 degrees, required two flight plan changes from AOP by design. First, SARA received an RTA that forced it to slow down (a "speed advisory" not involving PBGA). This caused OPV, which was following a parallel route (nearly the same route, but with an offset for safety during the test), to begin to overtake SARA. Eventually the time came when the overtaking conflict was within AOP's lookahead time, requiring AOP to use PBGA to attempt to resolve the conflict.

After the speed change for the new RTA, AOP was still predicting an on-time arrival at the RTA and predicted a cruise speed of about 80 kt CAS. After AOP detected the conflict with OPV, it started a rapid-fire sequence of PBGA requests (not waiting a full cycle), apparently due to "flicker" (instability in the prediction of the conflict).

Because of the geometry and constraints of the encounter, the only way for AOP to avoid a LOS was to speed up and stay ahead of OPV. Based on a partial examination of the data records from PBGA, it appears that any flight plan that arrived less than 17 seconds early at the RTA had a conflict with the OPV.

The fourth PBGA attempt for CR produced a resolution that was executed. In this resolution, the flight plan increased speed to about 96 kt CAS (according to AOP's prediction) and later would have decreased speed to 75 kt CAS (according to AOP's prediction), finally arriving about 17 seconds before the RTA.

After AOP executed the CR, causing AOP to predict that the aircraft would be early at the RTA, AOP made 18 requests to PBGA for RTA resolutions. The first request produced advisories that were not executed; most of the other requests apparently failed.

This run has been considered as one of the runs possibly affected by AOP's use of zero wind speed because AOP predicted the minimum allowed cruise speed, 75 kt CAS, for a portion of the final flight plan. The run ended before the aircraft reached that part of the flight plan, but if AOP had had a correct wind prediction it could have achieved a lower ground speed there without going below 75 kt CAS, and therefore it could have predicted a higher ground speed earlier in the flight plan while still arriving 17 seconds early at the RTA. But this does not seem to have any relevance to the metrics of the flight test.

By design, AOP will attempt to resolve the RTA if the active flight plan does not meet the RTA, but not if that would create a conflict. Since it was apparently not possible for AOP to meet the RTA while avoiding a conflict, the missed RTA (and the sequence of RTA resolutions due to it) would have occurred whether AOP had used zero wind speed or the predicted wind. There was therefore no wind effect on this run that could have affected the flight test results.

F.3 October 23

F.3.1 Maneuver M10a

In the run for maneuver M10a with 10 kt wind from 300 degrees, when a new RTA was received about a minute after T_0, a new 4D flight plan immediately lowered AOP's predicted cruise airspeed to 75 kt CAS. At that point AOP commenced a series of RTA resolutions that eventually called PBGA 28 times and stopped only when AOP detected a conflict with FLT0253, at which point PBGA started attempting CRs instead.

The RTA resolutions in this run were expected behavior and the pilot was instructed to observe but not execute them. The impact of wind on this run is that AOP predicted the arrival to be late by 160 seconds, which is presumably much later than the experiment called for, and therefore the resolutions are likely to be more extreme than they would have been with a correct wind prediction.

Impacts: Predicted to miss RTA by extreme amount; possible extreme resolution.

F.3.2 Maneuver M13a

In the run for maneuver M13a with 10 kt wind from 300 degrees, after the new RTA, the cruise speed increased to 109 kt CAS and remained so for the rest of the run, since no RTA resolutions were executed (although AOP computed many). AOP predicted it would be about 45 seconds late at the RTA.

Similarly to the run for M10a, the behavior was expected. The most likely effects of wind are that the predicted arrival at the RTA was probably not as late as expected and that at least some of the RTA resolution flight plans produced by PBGA probably would have caused SARA to fly at greater than 109 kt CAS if executed (although the AOP predictions were consistently below 109 kt CAS).

Impacts: The flight plan immediately after the RTA likely exceeded 109 kt CAS; likely predicted to miss RTA by less than scenario called for; some unexecuted resolution flight plans might have resulted in actual airspeeds greater than 109 kt CAS if executed.

F.3.3 Maneuver M13b

In the run for maneuver M13b with 10 kt wind from 300 degrees, after the new RTA, the cruise speed increased to 109 kt CAS. About 30 seconds later a hybrid resolution lowered the airspeed. This is expected behavior. The impact of wind on this run is presumably that the actual airspeed prior to the resolution was higher than AOP predicted due to the headwind, but it likely did not exceed SARA's guidance limits.

Impacts: The flight plan immediately after the RTA likely exceeded 109 kt CAS; likely predicted to miss RTA by less than scenario called for.

F.4 October 24

F.4.1 Maneuver M21

In the run for maneuver M21 with 20 kt wind from 240 degrees, AOP predicted a cruise airspeed of 75 kt CAS at all times. This caused AOP to generate two RTA resolutions in PBGA before the conflict and several more RTA resolutions afterward. None of these resolutions was executed. AOP called CR once and the resulting maneuver was executed.

In that run, AOP constantly predicted that it would arrive six or eight seconds early at the RTA. This likely meant that the initial flight plan AOP gave to SARA also crossed the path of OPV a little early, but if there was any difference it was not enough to prevent the desired conflict from being detected. The conflict location may have been slightly different than it would have been with a correct wind prediction.

Impacts: Undesired prediction of missed RTA; undesired RTA resolutions before conflict; possibly undesired RTA resolutions after CR; predicted conflict location may have shifted.

F.4.2 Maneuver M39

In the remaining runs on October 24 the predicted cruise speeds were all between 80 and 95 knots (which would have made it unlikely for AOP's use of zero wind speed to cause problems), except for M39 with 10 kt wind from 180 degrees. In that run the cruise speed was about 106 kt CAS but never exceeded 106.7 kt CAS. The predicted wind speed and direction would not likely have caused this airspeed to exceed 109 kt CAS when the 4D plan was flown.

F.5 October 25

F.5.1 Maneuver M42

In the run for maneuver M42 with 20 kt wind from 240 degrees, AOP predicted a cruise airspeed of 75 kt CAS at all times. AOP initially predicted it would arrive about 7 seconds early at the RTA, and it sent four RTA resolution requests to PBGA before it detected a conflict. After the conflict was resolved, AOP still predicted it would be about 4.5 seconds early at the RTA, but it did not attempt another RTA resolution. This suggests that the logic that triggers an RTA resolution was not using the same threshold that was specified for reporting that the aircraft is arriving too early or too late; if so, this would be a defect in AOP's use of the configuration. (AOP should have started an RTA resolution if it predicted it would arrive more than 1.5 seconds early at the RTA.)

Except for the RTA resolutions (which were ignored) and a possible shift in the location of the conflict (see October 24), there do not appear to be any impacts of wind on this run.

Impacts: Undesired prediction of missed RTA; undesired RTA resolutions before conflict; possibly undesired RTA resolutions after CR; predicted conflict location may have shifted.

F.5.2 Maneuver M43

In the run for maneuver M43 with 20 kt wind from 240 degrees, AOP initially predicted a cruise airspeed of 75 kt CAS and predicted it would arrive about 8 seconds early at the RTA. AOP attempted two RTA resolutions before the conflict was detected. After the conflict was resolved, AOP predicted it would arrive about 12 seconds early at the RTA. AOP attempted 12 more RTA resolutions.

As with other runs where AOP consistently predicted minimum airspeed, the flight plan may have produced a ground speed faster than expected and may have shifted the conflict location.

Although the run for M43 said to select and execute another resolution if a new conflict occurred, no new conflict was detected during this run. Instead, near the end of the run two of the RTA resolutions were executed. The first executed RTA resolution reduced the predicted RTA error to 6 seconds early, and the second executed RTA resolution eliminated the predicted error at the RTA. (The predicted arrival was within bounds and no error was reported.) It has not been determined whether either of these RTA resolutions would have occurred if AOP had used the predicted wind in its calculations.

Impacts: Undesired predictions of missed RTA; undesired RTA resolutions before conflict; possibly undesired RTA resolutions after CR; predicted conflict location may have shifted.

F.5.3 Maneuver M1

In the run for maneuver M1 in group 2 with 20 kt wind from 240 degrees, AOP initially predicted a cruise airspeed of 75 kt CAS and predicted it would arrive about seven or eight seconds early at the RTA. This caused AOP to request RTA resolutions from PBGA four times before the conflict with OPV was detected. After the conflict was resolved, AOP predicted a cruise CAS of 77 kt and an on-time arrival at the RTA.

Impacts: Undesired predictions of missed RTA; undesired RTA resolutions before conflict; predicted conflict location may have shifted.

F.5.4 Maneuver M2

In the run for maneuver M2 in group 1 with 20 kt wind from 240 degrees, during the entire run, AOP predicted a cruise airspeed of 75 kt CAS and predicted it would arrive about 11 or 12 seconds early at the RTA. This caused AOP to request RTA resolutions three times before the conflict with OPV was detected. After the conflict alert, all PBGA requests were for CR, as expected.

Impacts: Undesired predictions of missed RTA; undesired RTA resolutions before conflict; predicted conflict location may have shifted.

F.5.5 Maneuver M3

In the run for maneuver M3 in group 2 with 20 kt wind from 240 degrees, at the start of the run, AOP predicted a cruise airspeed of 75 kt CAS and predicted it would arrive about 12 seconds early at the RTA. AOP computed an RTA meet-time resolution and sent it to the display. About 15 seconds after the start of the run, AOP detected a conflict with OPV and started a CR cycle which produced one resolution, which was a lateral resolution. The resolution was then executed. After that time, AOP predicted a cruise speed of about 77.5 kt CAS and an on-time arrival.

Impacts: Undesired predictions of missed RTA; undesired RTA resolutions before conflict; predicted conflict location may have shifted.

F.5.6 Maneuver M4

In the run for maneuver M4 in group 1 with 20 kt wind from 240 degrees, during the entire run, AOP predicted a cruise airspeed of 75 kt CAS and predicted it would arrive about 18 or 19 seconds early at the RTA. This caused AOP to request RTA resolutions six times before the conflict with OPV was detected. After the conflict alert, all PBGA requests were for CR, as expected.

Impacts: Undesired predictions of missed RTA; undesired RTA resolutions before conflict; predicted conflict location may have shifted.

F.5.7 Maximum airspeed

There were no runs on October 25 where AOP's predicted cruise CAS was near the maximum.

F.6 October 26

AOP predicted cruise airspeeds between 79 and 80 kt CAS at all times in all runs on this test day. The relatively low predicted airspeed was probably due to using T_0 locations selected for the predicted wind of 20 kt from 300 degrees, but no evidence was found of any effect of AOP's use of zero wind speed on the behaviors measured in this report.

F.7 Summary

The most obvious effect on the metrics is that it is necessary to exclude RTA resolutions that likely would not have occurred if AOP had used the predicted wind in its calculations.

Appendix G: Priority Rules

The purpose of priority rules in AOP is to reduce the workload on operators by designating one of the aircraft in any conflict between two aircraft to be the aircraft that must resolve the conflict. That is, priority rules give one of the aircraft in the conflict the "right of way," meaning it does not have to change its flight plan immediately, while the aircraft without the "right of way" is required to change its plan. This also help avoids the possibility that both aircraft concurrently make changes in their flight plans and that the new flight plans result in a new conflict with a shorter TTFLOS, increasing workload and decreasing the ability of both aircraft to resolve all conflicts strategically.

Several choices of priority rules are implemented in AOP. The priority rules configured for AOP during the flight test were the "HITL" priority rules, one of several sets of priority rules that function by evaluating multiple individual rules one at a time. Each individual rule examines data from the ownship and traffic aircraft and returns one of the three results "OWNSHIP moves" (indicating that the ownship must maneuver), "TRAFFIC moves" (indicating that the traffic aircraft must maneuver), or "INCONCLUSIVE" (if this rule cannot determine which aircraft must maneuver). The individual rules are applied in a fixed sequence until a rule is found that determines that either the ownship or the traffic aircraft must maneuver. The result of that rule then determines the right of way. If no individual rule is found that determines which aircraft must maneuver, the ownship is required to maneuver.

G.1 Equipage priority rule

The equipage priority rule examines whether the "flight mode" of each aircraft is "autonomous" (using AOP to resolve conflicts) or "managed" (flying under conventional flight rules). The result of this rule is then given by Table G1.

Table G1. Calculation of the equipage priority rule.

		Traffic flight mode	
		managed	autonomous
Ownship flight mode	managed	INCONCLUSIVE	TRAFFIC moves
	autonomous	OWNSHIP moves	INCONCLUSIVE

G.2 Mixed operations priority rule

In the mixed operations priority rule, if the traffic flight mode is "managed" then the OWNSHIP must move. This result of the mixed operations priority rule can be enforced by the equipage priority rule in practice.

If the traffic flight mode is "autonomous," the mixed operations priority rule examines whether each aircraft's guidance mode is "strategic" or "tactical." The result of this rule is then given by Table G2.

Table G2. Calculation of the mixed operations priority rule when the traffic aircraft is "autonomous."

		Traffic guidance mode	
		strategic	tactical
Ownship guidance mode	strategic	INCONCLUSIVE	TRAFFIC moves
	tactical	OWNSHIP moves	INCONCLUSIVE

G.3 Vertical priority rule

The vertical priority rule attempts to determine whether each aircraft will be climbing, descending, or flying level at the time of first LOS. This determination is made by examining the predicted vertical rate of each aircraft at the time of first LOS. The result of this rule is then given by the Table G3.

Table G3. Calculation of the vertical priority rule.

		Traffic vertical speed, feet/minute		
		< -150	between -150 and 150 inclusive	> 150
Ownship vertical speed, feet/minute	< -150	INCONCLUSIVE	TRAFFIC moves	TRAFFIC moves
	between -150 and 150 inclusive	OWNSHIP moves	INCONCLUSIVE	TRAFFIC moves
	> 150	OWNSHIP moves	OWNSHIP moves	INCONCLUSIVE

G.4 Horizontal priority rule

The horizontal priority rule is based on the predicted true track angle of each aircraft at the time of first LOS, and, in some cases, on the predicted ground speed of each aircraft at the time of first LOS.

The rule is applied as follows. First, the predicted true track angle in degrees of the ownship at first LOS is rounded to the nearest integer. The value $\mathrm{ownship\_track\_degrees}$ is defined as an integer directional angle (in degrees) equivalent to the rounded true track angle of the ownship (the same number or differing by a multiple of 360) so that $-179 \leq \mathrm{ownship\_track\_degrees} \leq 180$. Similarly, the predicted true track angle in degrees of the traffic aircraft at first LOS is rounded to the nearest integer and the value $\mathrm{traffic\_track\_degrees}$ is defined as an integer directional angle (in degrees) equivalent to the rounded true track angle of the traffic aircraft so that $-179 \leq \mathrm{traffic\_track\_degrees} \leq 180$.

The value $\mathrm{delta\_track}$ is defined as the absolute value of the difference between $\mathrm{ownship\_track\_degrees}$ and $\mathrm{traffic\_track\_degrees}$. Therefore $\mathrm{delta\_track}$ is an integer between 0 and 359 inclusive.

The cases where $\mathrm{delta\_track} \leq 5$ or $\mathrm{delta\_track} \geq 355$ are considered "overtaking" cases, that is, both aircraft are going in approximately the same direction. (The case $\mathrm{delta\_track} \geq 355$ occurs when one of the two angles is near 180, the other is near -179, and the two angles are equivalent to angles that differ by 5 degrees or less.) In these cases the aircraft with the lesser ground speed has priority and the aircraft with the greater ground speed "must move." That is, if the ownship's ground speed is greater than the traffic aircraft's, the result is "OWNSHIP moves"; otherwise the result is "TRAFFIC moves." (The case where the ground speeds of the two aircraft in an "overtaking" conflict are exactly equal is unlikely to occur in practice, but in a future version of AOP it is expected to produce an "INCONCLUSIVE" result.)

The case where $175 \leq \mathrm{delta\_track} \leq 185$ is considered the "head on" case. That is, one aircraft's true track angle is the reciprocal (or nearly the reciprocal) of the other aircraft's true track angle. In the "head on" case, if any of the three conditions

- $\mathrm{traffic\_track\_degrees} < 0$ and $-5 < \mathrm{ownship\_track\_degrees} < 180$, or
- $\mathrm{traffic\_track\_degrees} \geq 175$ and $\mathrm{ownship\_track\_degrees} \geq 0$, or
- $\mathrm{traffic\_track\_degrees} = 180$

is true, the result is "OWNSHIP moves"; but if any of the three conditions

- $\text{ownship\_track\_degrees} < 0$ and $-5 < \text{traffic\_track\_degrees} < 180$, or
- $\text{ownship\_track\_degrees} \geq 175$ and $\text{traffic\_track\_degrees} \geq 0$, or
- $\text{ownship\_track\_degrees} = 180$

is true, the result is "TRAFFIC moves." These conditions cover every possible pair of track angles in the "head on" case.

If the two aircraft are neither "overtaking" nor "head on," the result is "OWNSHIP moves" if $\text{ownship\_track\_degrees} > \text{traffic\_track\_degrees}$; the result is "TRAFFIC moves" if $\text{traffic\_track\_degrees} > \text{ownship\_track\_degrees}$. (It is not possible for the two track angles to be equal in this case.)

Table G4 shows the result of the horizontal priority rule for all pairs of true track angles of the ownship and traffic aircraft.

G.5 Sequence of application of priority rules

The "HITL" priority rules used in the flight test (also known as the "MOES AFR" priority rules) apply the following individual rules in the sequence shown below:

- Equipage priority rule,
- Mixed operations priority rule,
- Vertical priority rule,
- Horizontal priority rule.

In the flight test, the Ownship (SARA) was considered the only "AOP-equipped" aircraft in the airspace (as discussed in Section B.1). All other aircraft were considered "managed," and therefore the result of the equipage priority rule was always "OWNSHIP moves," requiring AOP aboard the Ownship to attempt to resolve the conflict.

Two other sets of priority rules implemented in AOP (but not used in the flight test) are based on other sequences of the individual rules. The "MOES IFR" priority rules apply the following individual rules in the sequence shown below:

- Vertical priority rule,
- Horizontal priority rule.

The "Build 3" priority rules apply the following individual rules in the sequence shown below:

- Equipage priority rule,
- Vertical priority rule,
- Horizontal priority rule.

It is also possible to configure AOP to use "no" priority rules, that is, the AOP-equipped aircraft will conclude "OWNSHIP moves" for every conflict it detects.

The use of priority rules is based on the assumption that all AOP-equipped aircraft in the same airspace use the same priority rules, which are designed to force at least one aircraft in each conflict to maneuver by concluding "OWNSHIP moves" in the instance of AOP used by at least one of the aircraft. The ability to configure AOP to use alternative sets of priority rules is intended to support experimental investigation of the results of choosing one of these sets of priority rules to apply in a particular airspace.

Table G4. Results of the horizontal priority rule for each pair of ownship and traffic aircraft true track angles. Track angles must first be rounded to the nearest whole number and converted to an equivalent directional angle in the range from -179 to 180 inclusive. Cases where the aircraft with the greater ground speed "must move" are designated by the letter F (for "FASTER aircraft moves") and are color-coded yellow for easier recognition. In all other cases, either the result "OWNSHIP moves" (designated by the letter O, color-coded red) or the result "TRAFFIC moves" (designated by the letter T, color-coded green) can be determined merely by considering the predicted true track angles of both aircraft.